Inventory and Monitoring Program

National Park Service
U.S. Department of the Interior

DATA MANAGEMENT PLAN: HEARTLAND I&M NETWORK AND PRAIRIE CLUSTER PROTOTYPE MONITORING PROGRAM

Gareth A. Rowell, Michael H. Williams,
and Michael D. DeBacker

Heartland I&M Network and Prairie
Cluster Prototype Monitoring Program
Wilson's Creek National Battlefield
Republic, Missouri

March 8, 2005

Executive Summary

To carry out its mission, the National Park Service (NPS) serves as a steward for the preservation of America's national parks and their resources. The NPS has initiated a service-wide, network-based inventory and monitoring (I&M) program to address the need for science-based natural resource information available to parks. The purpose of the program is to design and implement long-term ecological monitoring and provide information for park managers to evaluate the integrity of park ecosystems and better understand ecosystem processes.

The NPS is implementing park inventories and vital signs monitoring programs in approximately 270 natural resource parks. The NPS has organized these parks into 32 networks, linked by geography and shared natural resource characteristics. Parks in each network will share resources and professional expertise to implement a core program of inventory and monitoring that focuses on the key indicators of change or “vital signs” identified for the network’s parks.

The Heartland I&M Network and Prairie Cluster Prototype Monitoring Program (HTLN) is composed of 15 parks in eight Midwestern states (Arkansas, Indiana, Iowa, Kansas, Minnesota, Missouri, Nebraska, Ohio) representing short/tallgrass prairie, Ozark highlands, and eastern deciduous forest eco-regions. Parks encompass a geographic area of over 720,000 square miles, include approximately 245,000 acres, and span roughly 2,700 feet of vertical relief.

This Data Management Plan identifies key data resources and processes to manage both inventory and monitoring data. A goal of this Plan is to describe the management of long-term monitoring data. Another goal of this Plan is to describe the resources and processes used to ensure high quality data. Specific objectives of the plan are:

- To describe the data management process;
- To describe roles and responsibilities of network staff and cooperators for managing data;
- To describe the current hardware and software environment in which we manage data;
- To summarize data and associated metadata collected and/or managed by the Heartland I&M Network and Prairie Cluster Prototype Monitoring Program;
- To schedule routine summary reports and trend analysis reports;
- To define future direction of data management activities in a work plan.

Network I&M staff and their cooperators make thousands of observations each year about plant and animal populations, communities and their environments. Taken together, these observations form a statistical representation of our sampling universe. In essence, the purpose of data management is to ensure that an accurate, complete record of those observations be maintained in perpetuity. Specific data management themes follow: data management infrastructure, data integration and exploration, and long-term data integrity and security. The focus of data management efforts within the Network is strategic within the context of service-wide data management activities. The service-wide Inventory and Monitoring Program has invested considerable resources in developing tools to archive and disseminate data (e.g. NPSpecies, Dataset Catalog, NatureBib). Further general guidance regarding data management procedures for handling and validating data is contained in the Data Management Plan: Prairie Cluster Prototype LTEM Program (DeBacker et al. 2002) and the Draft Data Management Protocol (Tessler & Gregson 1997). The Network data management system builds on

these resources while emphasizing and expanding the roles of geospatial database development and management during data collection and handling.

Assuring and maintaining data integrity is fundamental to the mission of an inventory and monitoring program and requires a considerable investment of staff time. The Data Manager, Project Manager, and Project Cooperator(s) share responsibility for data management within the Network. The Data Manager serves as GIS Specialist and Data Manager for the Network; the Project Manager supervises individual projects; the Project Cooperator carries out specific projects for the Network. Many of the Project Manager's duties involve close interaction with the Data Manager and Project Cooperator(s). In addition, the project manager will be the project Contracting Officers Technical Representative and/or designate a park Point-of-Contact to coordinate the project on behalf of the cooperating entity. Cooperating researchers have played an especially significant role in inventory projects and their associated data acquisition and data development. The cooperator Point-Of-Contact should be a permanent faculty member or agency employee who is responsible for the final delivery of data products.

In the interest of providing a more complete picture of our current project operations, we provide a list of hardware and software resources necessary to accomplish the goals and objectives as stated above. The automation equipment consists of approximately 10 Dell PC workstations and 1 mid-range dual CPU server. There has been an emphasis on Microsoft Access as software for database management. It is widely available, inexpensive, and interfaces with ESRI ArcGIS software to create geodatabases. Finally, we use SAS and NCSS for statistical analysis of monitoring data.

Data are direct measures of the population, community, or resource that is the focus of the protocol (for example, species inventory occurrence and abundance, rare species monitoring). Inventory data serves as the basis for future vital sign monitoring by identifying the fauna and flora. Monitoring data represent the change in vital signs over time. These vital signs follow sampling requirements defined in the monitoring protocol. For each Network, the project database is actually composed of many separate yet complimentary data sets. Each database should have event data and location data to reference observations in a specific time and place (i.e Natural Resource Database Template). Supplementary/ancillary environmental data assists in characterizing the specific environment or habitat that is associated with the primary vital sign. Metadata, developed as part of the data development process, also undergoes independent review to ensure compliance with NPS standards. Metadata is distributed at the NR-GIS clearinghouse with the I&M data sets, thus ensuring long-term value of the data. Additional information regarding GIS and metadata requirements, and metadata tools obtained at the NPS Midwest Region GIS website.

Almost all inventory and monitoring databases include a spatial component. The spatial data are usually represented as point data but may also be represented as line features (as in the case of transects) or polygons (for example, delineating habitat extents). ArcGIS personal geodatabases can be used as containers for both spatial and non-spatial data. The ArcGIS personal geodatabase allows shape files and coverages, representing GPS monitoring data, to be converted into feature classes. A feature class is a spatial data layer (point, line or polygon) stored in a relational database.

Data management procedures follow five key steps: acquisition, verification, validation, analysis and dissemination. In addition, storage, maintenance and security issues apply to all stages of the data flow. Data acquisition of Network I&M databases may include acquisitions through cooperative agreements with government agencies such as USGS-BRD or regional universities. Development of the Network natural resource knowledge base involved an initial phase of data mining of existing documents and reports. Data mining included bibliographic lists of reports about monitoring conducted within and around each park. Additional data mining captured park related monitoring, especially that conducted by neighboring agencies. These reports were scanned and optical character recognition of legacy documents was completed and then word-indexed. A database then summarized available information about monitoring activities within and around the Network parks.

Data acquisition may also be the direct result of field studies conducted by Network staff. Data acquisition may require data conversion from analog (usually paper) to digital or, on other occasions, be collected in digital form (GPS units and data loggers). Cooperators should obtain NPS park permits prior to initiating I&M studies, and both NPS staff and cooperators should obtain state and Federal fish and wildlife scientific permit letters prior to monitoring state or federally-listed species. All parties involved with any Network I&M project must be aware of any compliance issues regarding NPS Director's Orders.

Data entry is the initial set of operations in which raw data from paper field forms or field notebooks are transcribed into a computerized form (i.e., within a database). Data entry commences as soon as data collection is complete. Data entry forms and QA/QC features minimize error. Data entry forms reduce transcription errors through pick lists and value limits and provide controlled access to the database. Forms also control the sequence of data entry.

Data verification immediately follows data entry and involves checking the accuracy of computerized records against the original source—usually paper field records. While the goal of data entry is to achieve 100% correct entries, this is rarely accomplished. To minimize transcription errors, our policy is to verify all records to their original source by permanent staff. Further, 10% of records are reviewed a second time by the Project Manager and the results of that comparison reported with the data. The entire data set is re-verified if any errors are found. Once the computerized data are verified as accurately reflecting the original field data, the paper forms are scanned, archived and the electronic version is used for all subsequent data activities.

The process of reviewing computerized data for range and logic errors is the validation stage. Although data may be correctly transcribed from the original field forms, they may not be accurate or logical (eg. a stream pH of 25.0 or a temperature of 95°C is illogical). Certain components of data validation are built into data entry forms (e.g. range limits). Additional data validation can be accomplished during verification, if the operator is sufficiently knowledgeable about the data. The Project Manager will validate the data after verification is complete. Validation procedures seek to identify generic errors (e.g. missing, mismatched or duplicate records) as well as errors specific to particular projects. For example, validation of plant community data includes database query and comparison of data among years. The Project Manager must assure consistency between field forms and the database by noting how and why any changes were made to the data on the original field forms. In general, changes are made to field forms through marginal notes or attached explanations, not erasures. Once validation is complete, the data set is turned over to the Data Manager for archiving and storage.

Secure data archiving is essential for protecting data files from corruption. Once a data set has passed the QA/QC procedures specified in the protocol, a formal entry is made in the I&M Dataset Catalog. Project Managers maintain the current data files for their project. These data files are stored on the program server. The server uses a RAID 5 configuration to ensure data integrity. Access to these files is controlled through the use of assigned user privileges. Typically, the Data Manager has universal access to data sets, while access for Project Managers is limited to their projects. Prior to storage on the server all files are scanned for viruses. Additional security is provided through tape backups. A complete server dump to tape is completed at least two times per week and three tape copies are stored locally and four off-site. Tape copies are rotated as needed.

The process of data analysis follows, but is not limited to, five basic steps: data determination, exploratory data analysis, the statistical test, a posteriori power analysis, and interpretation of the results. Data determination is comprised of several questions: Are data discrete (i.e. nominal or ordinal) or continuous (i.e. interval or ratio)? How were data collected? How many groups? And most importantly, what is the question being asked of the data? Proper use of data analysis will allow the ability to detect a change in the monitoring program if a change has occurred and ensure long-term viability of the monitoring program.

Efficient reporting is important to encourage the use of monitoring data in management decisions. To promote efficient reporting, data management efforts during the summary and analysis phase focus on automation of routine reports. Following the appropriate review process, dissemination of data, reports, and other items (photos, sound recordings, etc.) will be in accordance with current NPS standards. Species lists and voucher data will be input to NPSpecies and ANCS+ and subsequently updated (if required), reports documented in NatureBib, and other natural resource data and GIS data forwarded to the Biodiversity Data Store and Natural Resource GIS Data Store. Network geodatabases will be forwarded to the I&M protocol website for distribution following QA/QC and metadata review. In addition, completion and review of monitoring protocols and standard operating procedures (SOPs) will, most likely, be required prior to data distribution.

All data collected are public property and subject to requests under the Freedom of Information Act (FOIA). The NPS policy regarding FOIA is discussed in Director’s Order 66 (which is in draft form at this time). Exemptions from FOIA are discussed in detail in the Director’s Order. Of specific concern to data managers of NPS inventory and monitoring data are: cave locations and resources, federally listed endangered or threatened animals and plants, and state listed endangered or threatened animals and plants. Certain cultural and water resource data are also exempt from FOIA so data managers need to become familiar with Director’s Order 66 prior to distributing information.

HTLN is just one of the 32 I&M networks in the National Park Service. Based on the data resources description in this document, the Network may be managing as many as 80 databases (including both inventory and monitoring data). The underlying goal of the program is to provide park managers with the scientific data they need to understand and manage park resources. It will be a challenge to meet this goal given the total number of databases. The WASO office has been developing the Vital Signs Framework in order to address these issues and bring a unifying structure to the I&M program (Oakley *et al.*, 2003; Fancy 2004).

Acknowledgements

This document is a revision of DeBacker, M. D., B. Witcher, and J. R. Boetsch (2002). Jennifer Haack provided new graphics for this revision. We wish to extend our thanks to Albert J. Rowell for biometric review in Appendix 1. We are in debt to each of the participants of the 2004 I&M Data Management Plan Workshop. Special thanks to all of the inventory and monitoring staff and partners who are making the Vital Signs Monitoring Program a reality in our National Parks.

Table of Contents

1. Introduction

1.1 Inventory and monitoring program

To carry out its mission, the National Park Service (NPS) serves as a steward for the preservation of America's national parks and their resources. The NPS has initiated a service-wide, network-based inventory and monitoring (I&M) program to address the need for science-based natural resource information available to parks. The purpose of the program is to design and implement long-term ecological monitoring and provide information for park managers to evaluate the integrity of park ecosystems and better understand ecosystem processes.

The NPS is implementing park vital signs monitoring programs in approximately 270 natural resource parks. The NPS has organized these parks into 32 networks, linked by geography and shared natural resource characteristics. Parks in each monitoring network will share resources and professional expertise to implement a core program of inventory and monitoring that focuses on the key indicators of change or “vital signs” identified for the network’s parks. The Heartland (HTLN) I&M Network and Prairie Cluster Prototype Monitoring Program is composed of 15 parks in eight Midwestern states representing tallgrass prairie, Ozark highlands, and eastern deciduous forest eco-regions (Figure 1). Ten of the fifteen network parks were established primarily to protect cultural and historic resources. The five largest parks in the network were established to protect outstanding natural resources.

The Network Parks include:

- Arkansas Post National Memorial (ARPO, 747 acres)
- Buffalo National River (BUFF, 94,293 acres)
- Cuyahoga Valley National Park (CUVA, 32,860 acres)
- Effigy Mounds National Monument* (EFMO, 2,526 acres)
- George Washington Carver National Monument (GWCA, 210 acres)
- Herbert Hoover National Historic Site (HEHO, 187 acres)
- Homestead National Monument of America* (HOME, 195 acres)
- Hopewell Culture National Historical Park (HOCU, 1170 acres)
- Hot Springs National Park (HOSP, 5,550 acres)
- Lincoln Boyhood National Memorial (LIBO, 200 acres)
- Ozark National Scenic Riverways (OZAR, 80,785 acres)
- Pea Ridge National Military Park (PERI, 4,300 acres)
- Pipestone National Monument* (PIPE, 282 acres)
- Tallgrass Prairie National Preserve* (TAPR, 10,894 acres)
- Wilson's Creek National Battlefield* (WICR, 1,750 acres)

Five parks in the network, which are also Prairie Cluster Prototype Program parks, are designated with an asterisk (*). Prairie Cluster Prototype parks benefit from established, on-going monitoring programs. An additional two parks monitored by the Prairie Cluster Prototype are part of the Northern Great Plains Network.

The parks in HTLN are located in eight Midwestern states (Arkansas, Indiana, Iowa, Kansas, Minnesota, Missouri, Nebraska, Ohio), encompass a geographic area of over 720,000 square miles, include approximately 235,000 acres, and span roughly 1,700 feet of vertical relief.

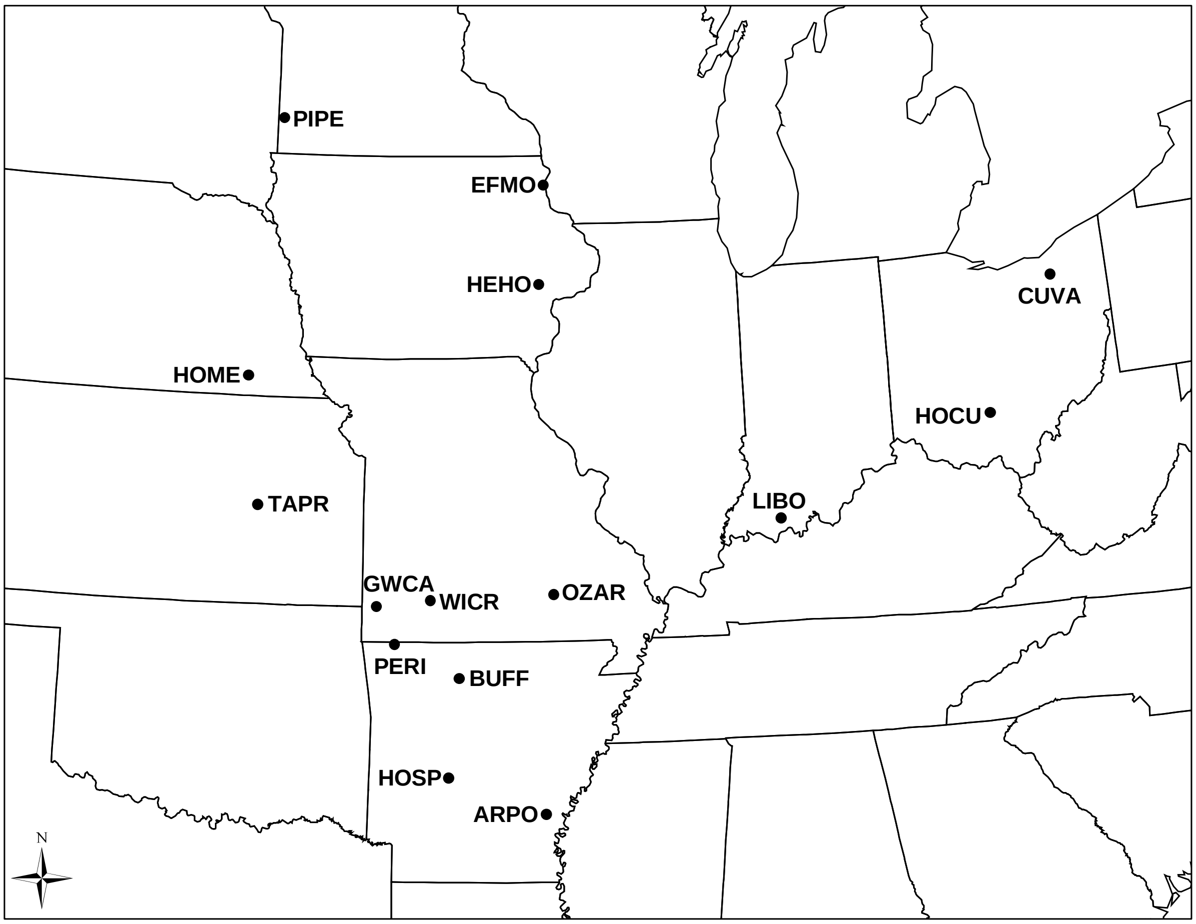


Figure 1. Parks located in the Prairie Cluster and Heartland Network.

1.2 Data management goals and objectives

Network staff and their cooperators make thousands of observations each year about plant and animal populations, communities and their environments. Taken together, these observations form a statistical representation of our sampling universe. In essence, the purpose of data management is to ensure that an accurate and complete record of those observations is maintained in perpetuity. Specific data management objectives follow three themes: data management infrastructure, data integration and exploration, and long-term data integrity and security.

Data infrastructure

- To design a data management system comprised of ArcGIS personal geodatabases and Microsoft Office Access databases that contain spatial and non-spatial data tables, data entry forms, and summary reports.

- To implement a standardized file organization system.

Data integration and exploration

- To design common fields and tables to ensure compatibility among data sets.
- To create a user interface to easily explore data relationships, alter the period or spatial scale of interest, and answer questions about the data.

Long-term data integrity and security

- To utilize redundant data storage devices.
- To archive data and maintain an edit log.
- To create and maintain spatial and tabular metadata.

The focus of data management efforts within the Network is strategic within the context of service-wide data management activities. The Service-wide Inventory and Monitoring Program has invested considerable resources in developing tools to archive and disseminate data (e.g. NPSpecies, Dataset Catalog, NatureBib). Further general guidance regarding data management procedures for handling and validating data is contained in the Data Management Plan: Prairie Cluster Prototype LTEM Program (DeBacker et al. 2002) and the Draft Data Management Protocol (Tessler & Gregson 1997). The Network data management system builds on these resources while emphasizing and expanding the roles of geospatial database development and management during data collection and handling.

This Data Management Plan identifies key data resources and processes to manage both inventory and monitoring data. A goal of this Plan is to describe the management of long-term monitoring data. Another goal of this Plan is to describe the resources and processes used to ensure high quality data. Specific objectives of the plan are:

- To describe the data management process;
- To describe roles and responsibilities of network staff and cooperators for managing data;
- To describe the current hardware and software environment in which we manage data;
- To summarize data and associated metadata collected and/or managed by the Network;
- To schedule routine summary reports and trend analysis reports;
- To define future direction of data management activities in a work plan.

While the Data Management Plan provides general data management guidance, project-specific data descriptions and procedures may also be found in the monitoring protocols. For each of the monitoring components (e.g. plant communities, grassland birds), the monitoring protocol details sample methods, summary routines and report format. Some of the protocols also include project-specific quality control procedures for conducting fieldwork. A database of I&M Program monitoring protocols is maintained at the national level by Dr. Steven Fancy at:
http://science.nature.nps.gov/im/monitor/protocoldb.cfm

2. Personnel Resources

Assuring and maintaining data integrity is fundamental to the mission of an inventory and monitoring program and requires a considerable investment of staff time. The Data Manager and Project Leader share responsibility for data management within the Network organization (Figure 2). Typically, the Project Leader is responsible for data collection, data entry, verification and validation, as well as data summary, analysis and reporting. The Data Manager is responsible for data archiving, security, dissemination and database design. Furthermore, the Data Manager, in collaboration with the Project Leader, develops data entry forms and other database features to assure quality assurance/quality control (QA/QC) and automate routine report generation. The Data Manager is ultimately responsible for developing adequate QA/QC procedures within the database management system and ensuring that appropriate data-handling procedures are followed. Initial inventory and monitoring work was conducted by either Network personnel or cooperating researchers (university or private researchers).

Table 1. HTLN staff resources directed towards data management.

Description	No. of Staff Positions	% of Time	Total FTE
Aquatic Resources Monitoring Leader	1	20	0.2
Botanist STF	1	30	0.3
Botanist	1	30	0.3
Data Manager	1	60	0.6
Fisheries Biologist	1	30	0.3
GIS Specialist	1	60	0.6
Plant Ecologist	1	30	0.3
Program Coordinator	1	10	0.1
Quantitative Ecologist	1	20	0.2
Seasonal Biotechnicians	2	30	0.6
Term Data Manager	1	60	0.6
Wildlife Ecologist	1	30	0.3
		Total FTE:	**4.4**
		Percent Total Program:	**31%**

Table 1 indicates staff resources directed towards data management. As the table shows, HTLN staff investment in data management is large, with approximately 31% of all FTE effort going into the management of I&M data. We estimate program managers (i.e., the Program Coordinator and Aquatic Resources Monitoring Leader) spend 10-20% of their time on data management. Staff botanists and ecologists contribute approximately 30% of their time towards this activity. Finally, we estimate data managers and GIS technical staff spend 60% of their time on data management.

2.1 Data Manager

The Network I&M Data Manager supervises a team of specialists who collectively have the responsibility to:

1) Write and maintain a Network Data Management Plan;

2) Work with project managers to ensure that data sets are fully documented and validated. This includes development of FGDC-compliant metadata;

3) Maintain archival copies of data sets and appropriate documentation;

4) Update data management aspects of project protocols in conjunction with project managers.

5) Integrate tabular data with spatial data in a GIS system. Currently this involves the development of personal geodatabases in ArcGIS. The resulting database integrates both spatial and non-spatial data in a single MS Access file;

6) Maintain and update those elements of the local area network relevant to data management. (LAN administration will be increasing shifted to regional IT specialists with the service-wide implementation of Windows 2003 Server);

7) Provide basic training in the use of GIS and database software and service-wide data management tools;

8) Contribute to NPS national office metadata and data clearinghouse efforts;

9) Contribute to regional and national discussions regarding data standards and integration/analysis issues.

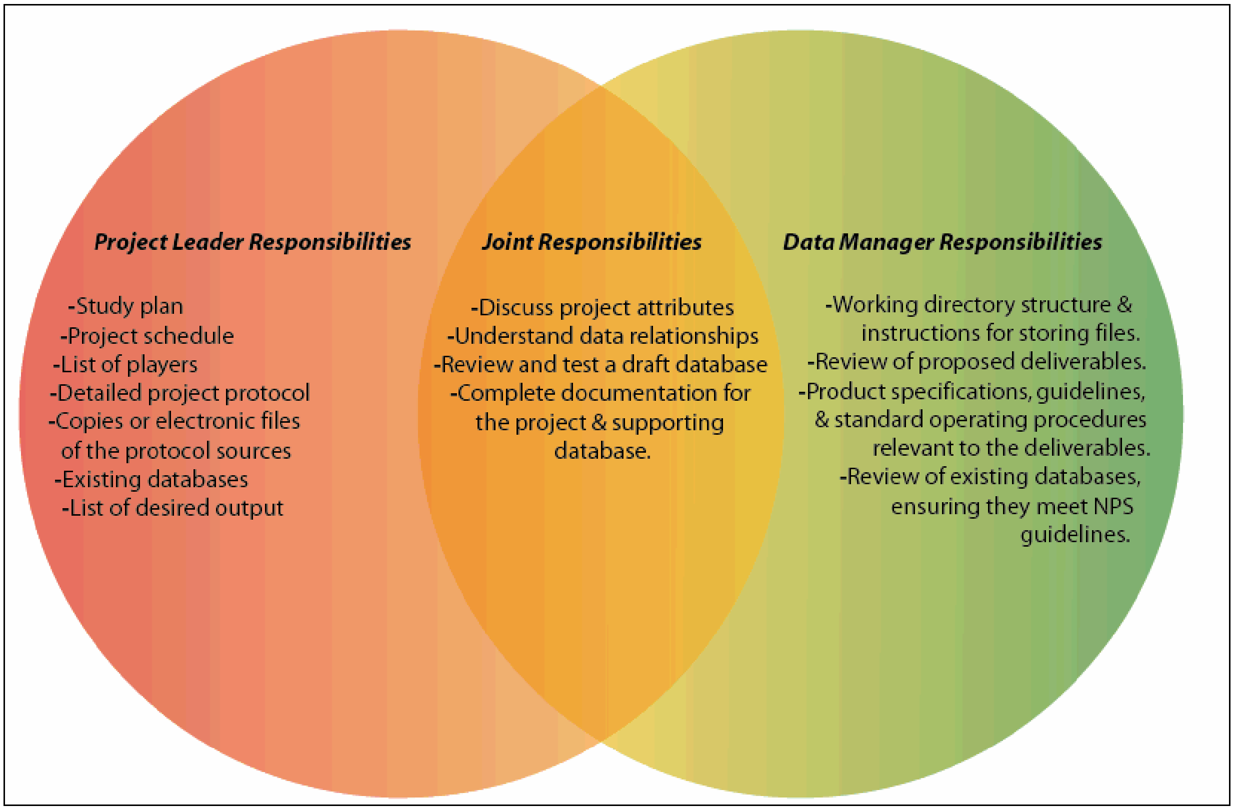


Figure 2. Responsibilities of the data manager and project leader (from CAKN Data Management Plan Draft, 2004).

2.2 Project Leader

The Network Project Leader supervises individual projects for the Network. As such, many of the Project Leader's duties involve close interaction with the Data Manager and Project Cooperator(s). The primary project management responsibilities of the position vary whether the project is done "in house" or via non-NPS entity and are to:

1) Supervise and certify all field operations, including staff training, equipment calibration and data collection.
2) Maintain concise explanatory documentation of all deviations from procedures defined in the monitoring protocols and standard operating procedures.
3) Supervise or perform data entry, verification and validation.
4) Work with Data Manager to fully document and maintain master data.
5) Maintain hard-copy files of data and ensure copies are stored in a second location. Create timely trip reports referencing important details of each field data collection period.
6) Coordinate changes to database structure and field data forms with the Data Manager.
7) Work with Data Manager to create general MS Access tools (queries, reports) for annual reports and for users to access their data.

8) Be the main point of contact concerning data content.

In addition, the Project Leader will be the project Contracting Officers Technical Representative and/or designate a park Point-of-Contact to coordinate the project on behalf of the cooperating entity. The cooperator Point-Of-Contact should be a permanent faculty member or agency employee who is responsible for the final delivery of data products.

2.3 Cooperators

Cooperating researchers have played an especially significant role in I&M projects and their associated data acquisition and data development. With regards to data management, there are some responsibilities which may require additional attention in order for the Data Manager – Project Leader - Cooperator relationship to work, specifically, in the areas of communication, project scope, and data quality control and assurance.

In addition to the Project Leader duties described above, the cooperator should:

1) Communicate with the Project Leader and park Point-of-Contact during the course of the project. Communication is especially helpful when significant project milestones have been met, or when unexpected complications arise.

2) Deliver data products that meet the specifications for that project. Delivery of final data products will most likely require the interaction of the Cooperator and Data Manager to ensure that useful data products will be provided, including data format, accuracy, QA/QC procedures, and the extent of analysis and reporting, if any. Cooperators should provide metadata that meet FGDC and other NPS standards with each data product.

3. Computer Resources

The utility of producing a list of Network computer resources may seem questionable given that most of these resources are quickly outdated and replaced by newer technology. However, in the interest of providing a more complete picture of our current project operations, we provide a list of hardware and software resources necessary to accomplish the goals and objectives as stated at the beginning of this document.

3.1 Network, Hardware and Peripherals

Hardware used by the Network currently consists of:

Quantity	Description
1	Cisco 2600 router
1	HP Procurve Switch 2224
1	Larscom SplitT Multiplexor
1	Dell PowerEdge 2650 Server
10	Dell Workstations (Dell Optiplex and Precision Workstation Series)
4	Laptops
5	Trimble GPS units
3	Digital cameras
2	35 mm SLR cameras

[Network equipment supports additional servers and workstations used by park staff]

3.2 Software

Quantity	Description
1	Microsoft Windows Server 2003 (Standard Edition)
10	Microsoft Windows XP (Professional)
10	Microsoft Office XP includes Access 2002
4	ESRI ArcGIS 8.3
1	NCSS/PASS statistical software
1	SAS PC statistical software
5	NPSpecies Desktop application
5	NatureBib Desktop application
1	Adobe (Pagemaker, Photoshop, Acrobat)
1	Quicken 2000 Data software
1	Procite

3.3 Digital Data Formats

The primary digital data formats in use by our Network are ESRI ArcGIS vector, ERDAS and tiff raster, and Access database .mdb files which may include spatial and non-spatial attribute data. ArcGIS vector data formats may include shape files, ARCINFO coverages, or ArcGIS personal geodatabases (using Access .mdb files). The GIS interchange format of choice is the non-compressed .e00 file. This format is preferred over coverages or shape files because all spatial and attribute information is contained in a single file in ASCII text format. Raster data are managed as either Leica (formerly ERDAS) Imagine .img files or as geotiff files. Geotiff files are typically faster to work with in the ESRI ArcGIS environment. Finally, the bulk of our I&M

data are stored in relatively small Access files, on a project by project basis. GPS spatial information may be linked to Access databases (see Section 6.1 below). The majority of the Network photos are in either jpeg or bmp format. Anticipated sound recordings will be in mpeg format.

4. Data Resources

Network data sources come from a variety of I&M natural resource projects. In this section, we give an overview of data associated with program inventory and monitoring projects. We also examine database components derived from sample databases. Finally, we look at how these components are integrated with spatial information, usually derived from GPS, to create an ESRI ArcGIS geodatabase.

4.1 Inventory Data

Based on the Network Inventory Study Plan (Boetsch et al. 2000), projects were initiated in four fiscal years (FY01-04). These include twelve herpetofaunal, nine bat, eight fish, seven vascular plant, seven bird, eight exotic plant, six mammal, and two deer projects to be conducted in a total of 15 parks (Table 2).

Table 2. Network park specific inventories by fiscal year.

Park	Bats	Birds	Deer	Exotic Plants	Fish	Herp	Mammals	Vascular Plants
ARPO	2003	2002				2001		
BUFF	2002	2002			2002	2002		2001
CUVA	2002			2001/02		2003		
EFMO		2002						2003
GWCA				2001	2002	2001	2004	2002
HEHO	2003	2002		2003		2002	2004	
HOCU	2003	2003			2003	2002	2003	
HOME	2003		2003	2003		2002	2004	2002
HOSP				2002	2002	2001		2001
LIBO				2003				2001
OZAR		2002/03				2001		
PERI	2002			2001	2002			2001
PIPE					2002	2002		
TAPR	2003		2002		2002	2002	2004	
WICR	2002				2002	2001	2004	

For each I&M project, the project database is actually composed of many separate yet complimentary data sets (see Figure 3). Each may have a spatial and non-spatial (attribute) components. The following categories describe and assist in the organization of the various data sets managed by a long-term monitoring program. The I&M staff or cooperator(s) directly collect location and event data, monitoring data, and habitat/environmental data. Metadata are typically created at later stages of data development either by the project manager or cooperator, interacting with the data manager.

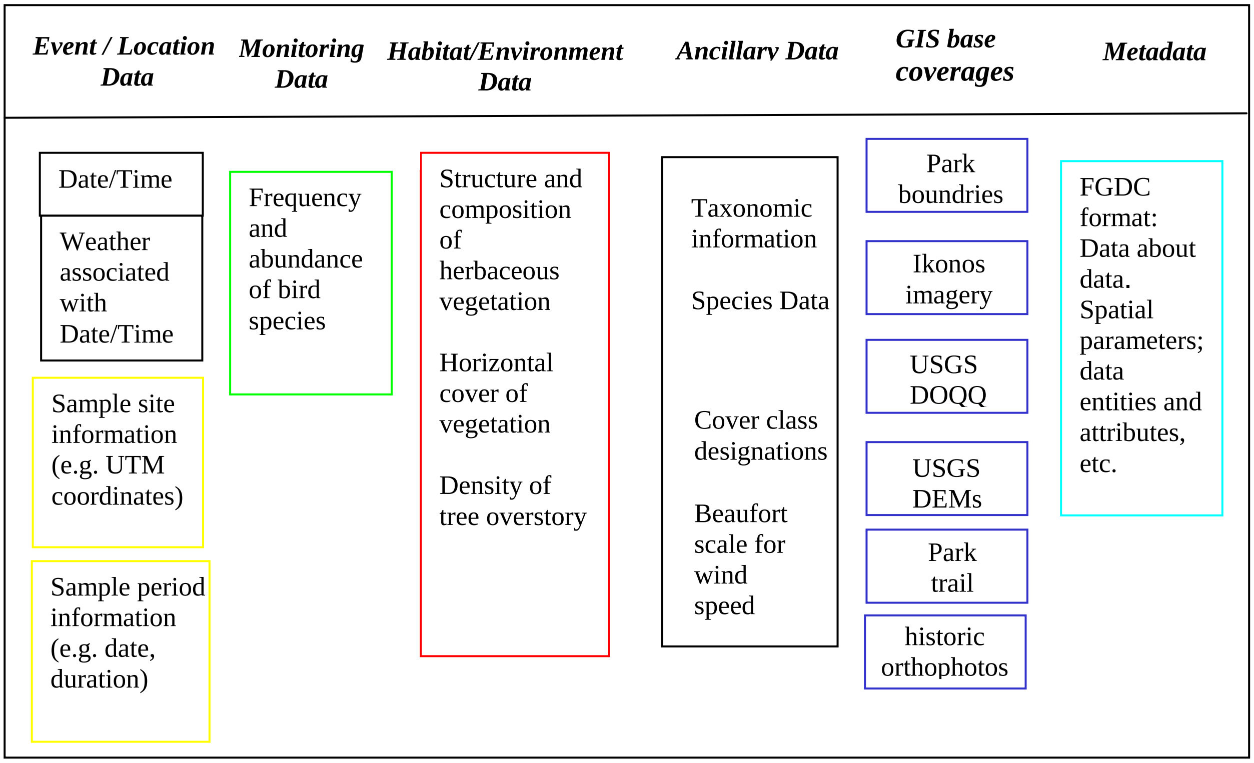


Figure 3. Relationships between data resources for an example monitoring project: grassland bird communities.

4.2 Monitoring Data

HTLN is an amalgamation of the Prairie Cluster Prototype Program and the Heartland Network. Several long-term monitoring projects initiated by the Prototype Program are now serving as guidance for the many new monitoring projects being undertaken by the network. The ongoing Prototype projects include: Missouri bladderpod, western prairie fringed orchid, prairie community structure, composition, and diversity, prairie stream fish communities (including Topeka shiner populations), adjacent land-use, aquatic macroinvertebrates, black-tailed prairie dog populations, and local climate. The new and ongoing projects for our Network are given in Table 3 (below). These projects include major projects to monitor forest plant communities, exotic forest and grassland species, stream macroinvertebrates, stream core elements, and a number of monitoring projects for individual species of concern.

4.3 Database Components

Event Data - Each monitoring database should have event data to track when the monitoring work is conducted. Typically, the event corresponds to the day and year of the observations. However, for some animal observations, time-of-day is critical. In such instances, hour and minute is recorded in addition to year, month, and day. This information is combined with park and project code to create a unique EventID which serves as an organizing field for the entire database.

Location Data – Together with event data, location data are necessary to place monitoring observations in a specific time and place. Location data are usually collected in UTM coordinates using mapping-grade GPS units. Differential correction of GPS positions may or may not be required, depending on the specifications of the monitoring protocol. A unique identifier called LocationID links the location data with all other data in the monitoring database. The LocationID consists of the park and project codes concatenated with a unique number.

Monitoring data – These are direct measures of the population, community, or resource that is the focus of the monitoring protocol (for example, plant species occurrence and abundance for plant community monitoring; prairie dog counts for prairie dog monitoring). Monitoring data represent the primary data for the project. Sampling requirements are defined in the monitoring protocol; they should be followed exactly.

Habitat / Environmental data – Supplementary environmental data may be collected in addition to the primary monitoring data. This is typically to assist in characterizing the specific environment or habitat that is associated with the primary vital sign. Examples might include water temperature, stream flow and substrate description for macroinvertebrates; or weather conditions and vegetation structure for bird monitoring.

Ancillary / Attribute data – Management actions (e.g. prescribed fire, herbicide application) and climate (e.g. precipitation) are examples of ancillary data that are applicable to all project areas. Attribute data, such as a taxonomic information, cover class designations, or other support information are generally not collected by staff, but instead are acquired from third parties.

GIS base coverages – Base coverages for parks generally include some or all of the following: park boundaries, USGS DEMs, hypsography, DRGs, quad boundaries, DOQQs, NPS park trails, roads, cultural sites, historic orthophotos, water quality monitoring sites, vegetation, burn units and fire management areas. Most geographic information system (GIS) datasets for the Prairie Cluster and Heartland Network can be obtained from the NPS GIS Clearinghouse at:
http://www.nps.gov/gis/data_info/clearinghouse.html

Metadata – The availability of standardized metadata for monitoring data sets is critical to the utility of the data sets. Metadata is developed as part of the data development process. It also undergoes independent review to ensure FGDC compliance. Metadata will then be distributed most likely at the clearinghouse with the I&M data sets, thus ensuring long-term value of the data. Additional information regarding GIS and metadata requirements, and metadata tools can be obtained at the NPS GIS Clearinghouse.

Voucher data
Digital data in the form of photographs and sound recordings i.e. data produced via recordings (birds, frogs, bats) representing taxa vouchers will be catalogued with the appropriate ANCS+ accession number, catalog number, and collection number and supporting metadata stored in NPSpecies.

When each of the above database components have undergone QA/QC procedures and the data verified, a geodatabase is developed. See section 4.4 for descriptions and examples of geodatabases.

Examples of I&M vital signs - monitoring projects are given in Table 3 below.

The relationships between Network monitoring projects and the Vital Signs Framework are described in Chapter 6 (below).

Table 3. List of Vital Signs selected for implementation in HTLN parks.

Level 1	Level 2	Level 3	Network Vital Sign Name	ARPO	BUFF	CUVA	EFMO	GWCA	HEHO	HOCU	HOME	HOSP	LIBO	OZAR	PERI	PIPE	TAPR	WICR
Air and Climate	Air Quality	Ozone	Ozone	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
		Wet and Dry Deposition	Wet and Dry Deposition	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
		Visibility and Particulate Matter	Visibility and Particulate Matter	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
		Air Contaminants	Air Contaminants	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
	Weather and Climate	Weather and Climate	Weather	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Geology And Soils	Geomorphology	Stream/ River Channel Characteristics	Fluvial Geomorphology		X									X				
			Stream Habitat/ Riparian Assessment								X					X		X
Water	Hydrology	Groundwater Dynamics	Spring Discharge*									X		X				
		Surface Water Dynamics	Stream Discharge	X	X	X	X	X	X	X	X	X		X	X	X	X	X
	Water Quality	Water Chemistry	Core Water Quality Parameters		X	X	X	X	X		X			X		X	X	X
			Drinking Water Quality*									X						
		Nutrient Dynamics	Nutrient Loading*		X	X	X							X		X		
		Toxics	Pathogens*		X	X	X							X		X		X
			Pollutant Metals			X								X				
		Aquatic Macroinvertebrates and Algae	Aquatic Macroinvertebrates—Prairie Streams					X			X					X		X
			Aquatic Macroinvertebrates—Rivers		X	X								X				
Biological Integrity	Invasive Species	Invasive/Exotic Plants	Exotic Forest Plants			X	X	X		X	X		X		X	X	X	X
		Invasive/Exotic Plants	Exotic Grassland Plants					X	X	X						X	X	X
		Invasive/Exotic Animals	Nutria*	X														
	Infestations and Disease	Insect Pests	Gypsy Moths*			X				X			X					
	Focal Species or Communities	Wetland Communities	Wetland Plant Communities			X												
		Grassland Communities	Prairie Community Structure, Composition, and Diversity				X	X	X		X					X	X	X

Table 3. List of Vital Signs selected for implementation in HTLN parks – continued.

Level 1	Level 2	Level 3	Network Vital Sign Name	ARPO	BUFF	CUVA	EFMO	GWCA	HEHO	HOCU	HOME	HOSP	LIBO	OZAR	PERI	PIPE	TAPR	WICR
Biological Integrity	Focal Species or Communities	Forest Vegetation	Forest Community Structure, Composition, and Diversity	X			X			X	X	X	X		X			X
		Fishes	Fish Community—Prairie Streams													X	X	
			Fish Community—Ozark Rivers		X									X				
		Birds	Landbirds						X	X							X	
		Mammals	Deer	X		X							X		X			X
			Gray Bat*		X									X				X
			Indiana Bat*		X	X								X				
			Ozark Big-Eared Bat*		X													
	At-risk Biota	Threatened and Endangered Species and Communities	American Alligator*	X														
			Missouri Bladderpod															X
			Ozark Hellbender											X				
			Topeka Shiner													X	X	
			Western Prairie Fringed Orchid													X		
Ecosystem Patterns and Processes	Land Cover and Use	Land Cover and Use	Land Cover/Land Use		X		X					X	X	X	X	X		X

Table 4. Implementation information for HTLN Vital Signs Monitoring

Vital Sign	Protocol Name	Attributes Measured	Protocol Developed by:	Protocol Implemented by:	NPS Contact
Ozone	NPS Air Atlas	Ozone	NPS Air Resources Division	NPS Air Resources Division	Mike DeBacker
Wet and Dry Deposition	NPS Air Atlas	Nitrates and sulfates	NPS Air Resources Division	NPS Air Resources Division	Mike DeBacker
Visibility and Particulate Matter	NPS Air Atlas	IMPROVE monitoring	NPS Air Resources Division	NPS Air Resources Division	Mike DeBacker
Air Contaminants	NPS Air Atlas	Persistent organic pollutants, fine particles, carbon dioxide, methane, UV-B	NPS Air Resources Division	NPS Air Resources Division	Mike DeBacker
Weather	1) Consolidation of Weather Service data and USGS Stream Flow Data 2) Weather Monitoring Protocol for Two Prairie Parks (PIPE, WICR)	1) Daily ambient temperature, precipitation, max. temp., min. temp., degree days 2) Temperature, RH, wind speed, wind direction, precipitation, solar radiation, soil temperature, soil moisture, snow depth, fuel moisture	1) National Weather Service and Atmospheric Science Program, University of Missouri – Columbia 2) Atmospheric Science Program, University of Missouri – Columbia and MO Field Station, NPWRC, Biological Resources Division, U.S. Geological Survey	1) National Weather Service and Atmospheric Science Program, University of Missouri – Columbia 2) Atmospheric Science Program, University of Missouri – Columbia	1) David Peitz 2) David Peitz
Fluvial Geomorphology	Physical Habitat Monitoring of Streams and Rivers	Channel longitudinal profile, cross section, channel planform, bank stability, sediment composition, photomonitoring	Southwest Missouri State University	NPS - HTLN	David Bowles
Stream Habitat/Riparian Assessment	Stream Habitat and Riparian Assessment for Prairie Streams (draft)	Bank erosion (stability), banks substrate type, bank height and slope, bank vegetation condition, riparian vegetation condition, substrate characteristics, embeddedness, woody debris, vegetation cover	NPS – Prairie Cluster Prototype	NPS - HTLN	David Peitz
Stream Discharge	1) Consolidation of Weather Service data and USGS Stream Flow Data 2) Stream Habitat and Riparian Assessment for Prairie Streams (draft)	Discharge	1) USGS and Atmospheric Science Program, University of Missouri – Columbia 2) NPS – Prairie Cluster Prototype	1) US Geological Survey and Atmospheric Science Program, University of Missouri – Columbia 2) NPS - HTLN	1) David Peitz 2) David Peitz
Core Water Quality Parameters	1) Stream habitat and Riparian Assessment for Prairie Streams (draft) 2) Fish Community Monitoring in Prairie Streams with Emphasis on Topeka Shiner (*Notropis topeka*) 3) Macroinvertebrate Biomonitoring Protocol for Four Prairie Parks 4) Macroinvertebrate Monitoring Protocol for Buffalo National River and Ozark National Scenic Riverways	For all protocols: temperature, conductivity, pH, dissolved oxygen, turbidity	1) NPS - Prairie Cluster Prototype 2) NPS - Prairie Cluster Prototype 3) MO Field Station, NPWRC, Biological Resources Division, U.S. Geological Survey 4) NPS – HTLN and BUFF	1) NPS - HTLN 2) NPS - HTLN 3) NPS – HTLN and park Resource Managers 4) NPS - HTLN	1) David Peitz 2) David Peitz 3) David Peitz 4) David Bowles
Pollutant Metals	Monitoring protocol for lead (Pb) and mercury (Hg) within the Ozark National Scenic Riverways, Missouri	Lead and mercury levels in water, sediments, and animal tissue	Columbia Environmental Research Center, US Geological Survey	TBD	David Bowles
Aquatic Macroinvertebrates—Prairie Streams	Macroinvertebrate Biomonitoring Protocol for Four Prairie Parks	Species richness and abundance, indices of biological integrity including: family biotic index, and EPT ratio	MO Field Station, NPWRC, Biological Resources Division, U.S. Geological Survey	NPS – HTLN and park Resource Managers	David Peitz
Aquatic Macroinvertebrates—Rivers	Macroinvertebrate Monitoring Protocol for Buffalo National River and Ozark National Scenic Riverways	Species richness and abundance, indices of biological integrity including: family biotic index, and EPT ratio	NPS – HTLN and BUFF	NPS - HTLN	David Bowles

Exotic Forest Plants	Invasive Non-native Plant Monitoring in HTLN Parks	Presence, abundance, distribution, rate of spread	NPS - HTLN	NPS - HTLN	Craig Young
Exotic Grassland Plants	Invasive Non-native Plant Monitoring in HTLN Parks	Presence, abundance, distribution, rate of spread	NPS - HTLN	NPS - HTLN	Craig Young
Wetland Plant Communities	Wetland Monitoring Protocol for Indicators of Ecosystem Health in the Cuyahoga Valley National Park	Water level, water chemistry, wetland vegetation, soils, litter decomposition, habitat variables	NPS – CUVA and University of Akron	TBD	Kevin Skerl
Prairie Community Structure, Composition, and Diversity	Vegetation Community Monitoring Protocol for HTLN	Species richness, diversity and abundance, frequency, vegetation structure, habitat characteristics, photomonitoring	NPS – Prairie Cluster Prototype, Nature's Keepers Services and MO Field Station, NPWRC, Biological Resources Division, U.S. Geological Survey	NPS - HTLN	Alicia Sasseen
Forest Community Structure, Composition, and Diversity	Vegetation Community Monitoring Protocol for HTLN	Species richness, diversity and abundance, frequency, vegetation structure, habitat characteristics, photomonitoring	NPS – HTLN, Nature's Keepers Services and MO Field Station, NPWRC, Biological Resources Division, U.S. Geological Survey	NPS - HTLN	Alicia Sasseen
Fish Community—Prairie Streams	Fish Community Monitoring in Prairie Streams with Emphasis on Topeka Shiner (*Notropis topeka*)	Species richness, diversity and abundance, species-habitat relationships	NPS - Prairie Cluster Prototype	NPS - HTLN	David Peitz
Fish Community—Ozark Rivers	Fish Community Sampling Protocols for Buffalo National River, Ozark National Scenic Riverways, and Other Park Service Units in the Ozark Highlands of Arkansas and Missouri	Species richness, diversity and abundance, indices of community structure and biotic integrity including: percent of individuals of tolerant species, percent of individuals of invertivores	Arkansas District, Water Resources Division, US Geological Survey	NPS - HTLN	David Bowles
Landbirds	Bird Monitoring Protocol for Agate Fossil Beds National Monument, Nebraska and Tallgrass Prairie National Preserve, Kansas	Species richness, diversity and abundance, proportion of sites occupied, habitat characteristics, species-habitat relationships	NPS - Prairie Cluster Prototype	NPS - HTLN	David Peitz
Deer	White-Tailed Deer Monitoring Protocol for Heartland I&M Network and Prairie Cluster Prototype Monitoring Program Parks	Abundance, density, distribution, impact on vegetation	NPS - HTLN	NPS - HTLN	David Peitz
Missouri Bladderpod	Missouri Bladderpod Monitoring Protocol for Wilson's Creek National Battlefield	Population size, density, reproduction, habitat characteristics, and plant-habitat relationships	Truman State University and MO Field Station, NPWRC, Biological Resources Division, U.S. Geological Survey	NPS - HTLN	Craig Young
Ozark Hellbender	Habitat mapping, baseline inventory, and development of a monitoring protocol for the Ozark Hellbender (*Cryptobranchus alleganiensis bishopi*) within the Ozark National Scenic Riverways	Population size and distribution	Center for Biodiversity, Illinois Natural History Survey	TBD	David Bowles
Topeka Shiner	Fish Community Monitoring in Prairie Streams with Emphasis on Topeka Shiner (*Notropis topeka*)	Population size and distribution, reproductive success, species-habitat relationships	NPS - Prairie Cluster Prototype	NPS - HTLN	David Peitz
Western Prairie Fringed Orchid	Western Prairie Fringed Orchid Monitoring Protocol for Pipestone National Monument	Population size, distribution, reproduction, and plant-habitat relationships	MO Field Station, NPWRC, Biological Resources Division, U.S. Geological Survey	NPS - HTLN	Craig Young
Land Cover/Land Use	Land Use, Land Cover Monitoring Protocol for HTLN	Area and distribution, fragmentation and connectivity, invasion corridors, patch characteristics, land use conversion rate, land use change analysis, census data	University of Arkansas at Monticello	TBD	Gareth Rowell

4.4 Using ArcGIS Geodatabases

Almost all inventory and monitoring databases include a spatial component. The spatial data are usually represented as point data but may also be represented as line features (as in the case of transects) or polygons (for example, delineating habitat extents). ArcGIS personal geodatabases can be used as containers for both spatial and non-spatial data. The ArcGIS personal geodatabase allows shape files and coverages, representing GPS monitoring data, to be converted into feature classes. A feature class is a spatial data layer (point, line or polygon) stored in a relational database. In the case of personal geodatabases, the spatial data are stored in Access tables (Figure 4). Spatial tables in Access are identified by their GDB_ and fcl prefixes. These tables are managed from ArcGIS, either using ArcCatalog or ArcMap. The spatial tables should not be edited from within Access.

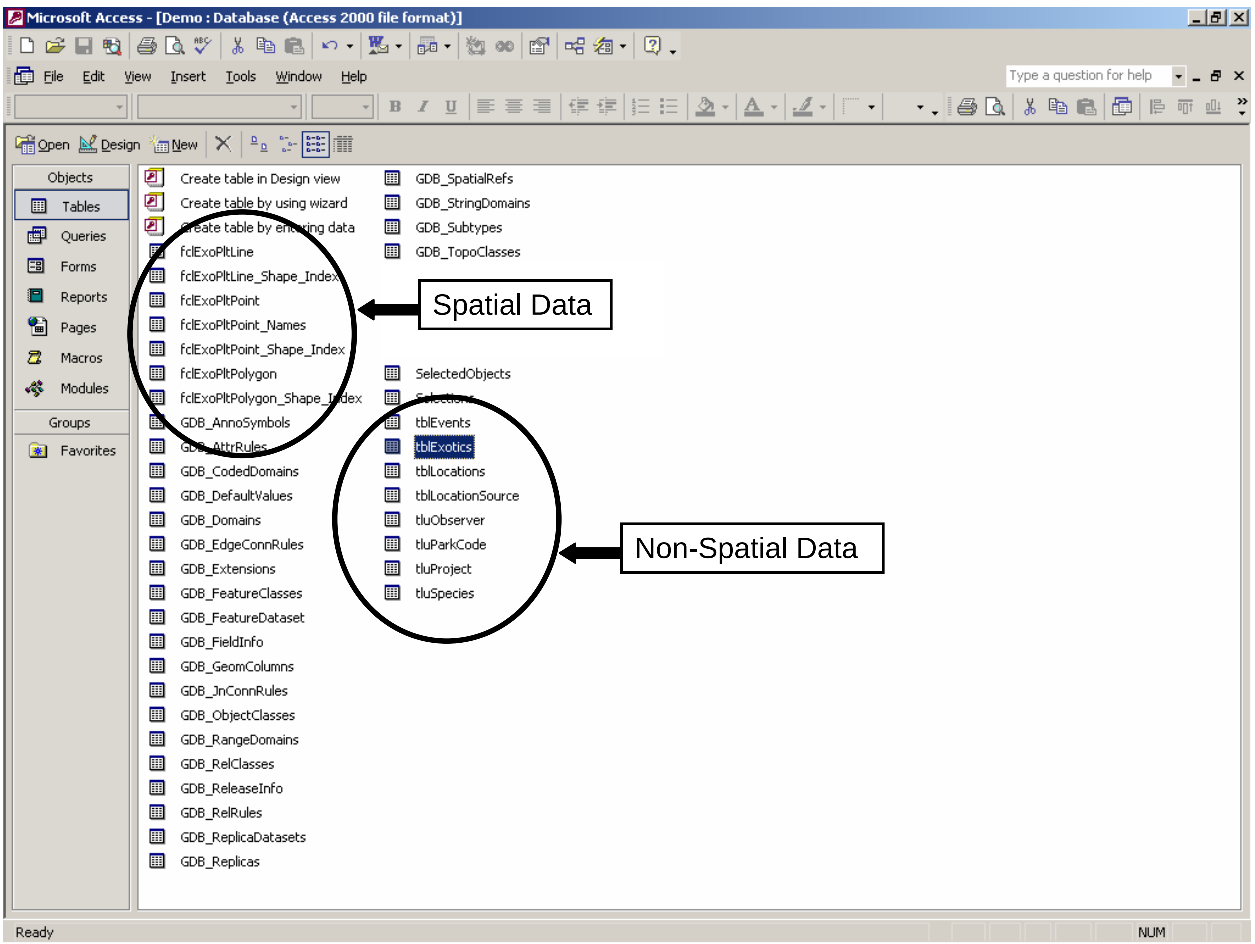


Figure 4. List of spatial (GDB_ & fcl) and attribute (tbl) tables in Access.

Most geodatabases in our program follow the same basic design (see Figure 4 and Figure 5). The feature class carries a linking field called LocationID which corresponds to the LocationID stored in the tblLocations table. Records from the inventory and monitoring databases tables can be joined to the feature class through this link. The beauty of this arrangement is that both spatial and non-spatial attribute data can be stored in a single Access file and viewed interactively in an ArcGIS environment (Figure 6). This makes small inventory and monitoring geodatabases extremely portable.

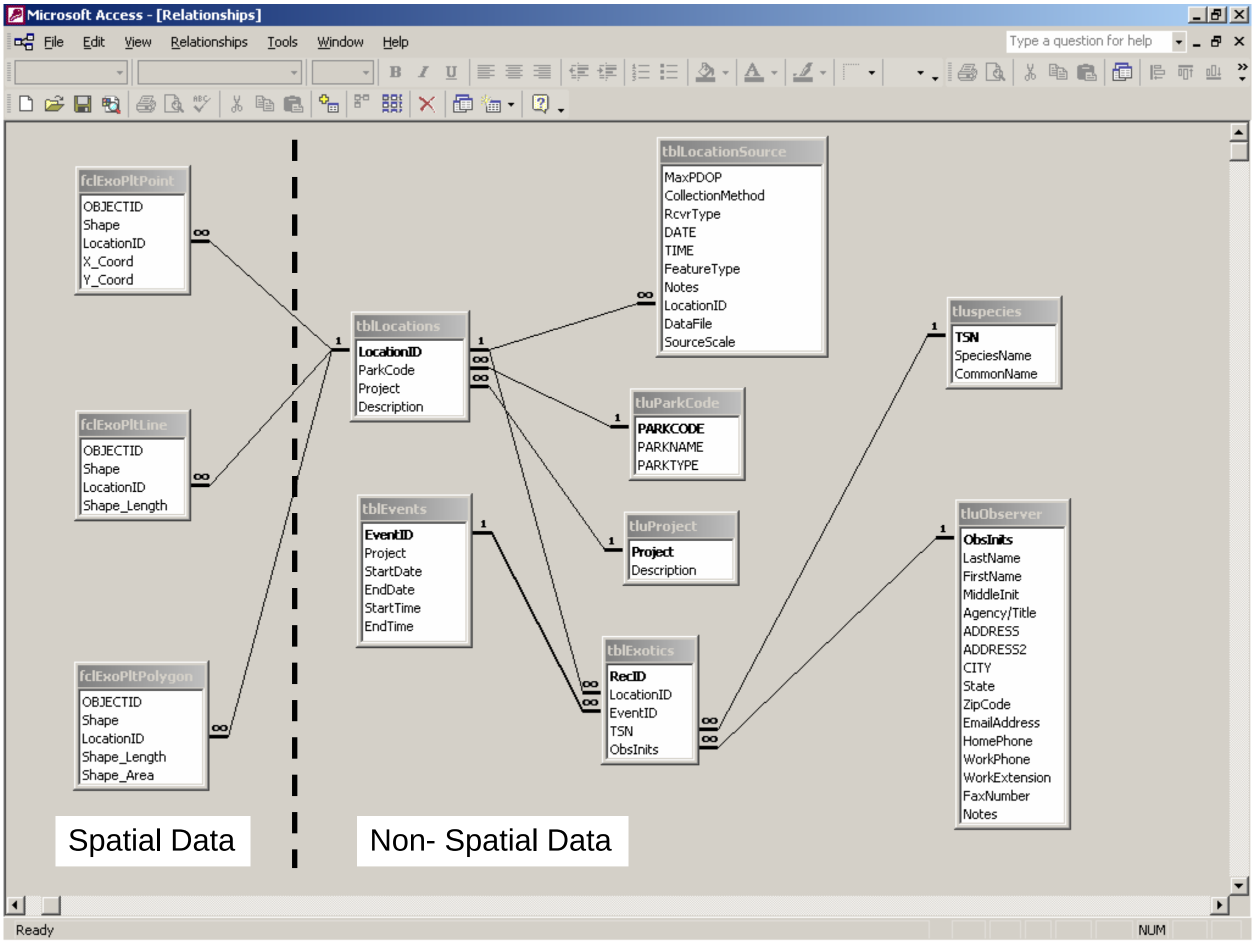


Figure 5. Standardized format of Geodatabase, Location and Period tables, and Lookup tables.

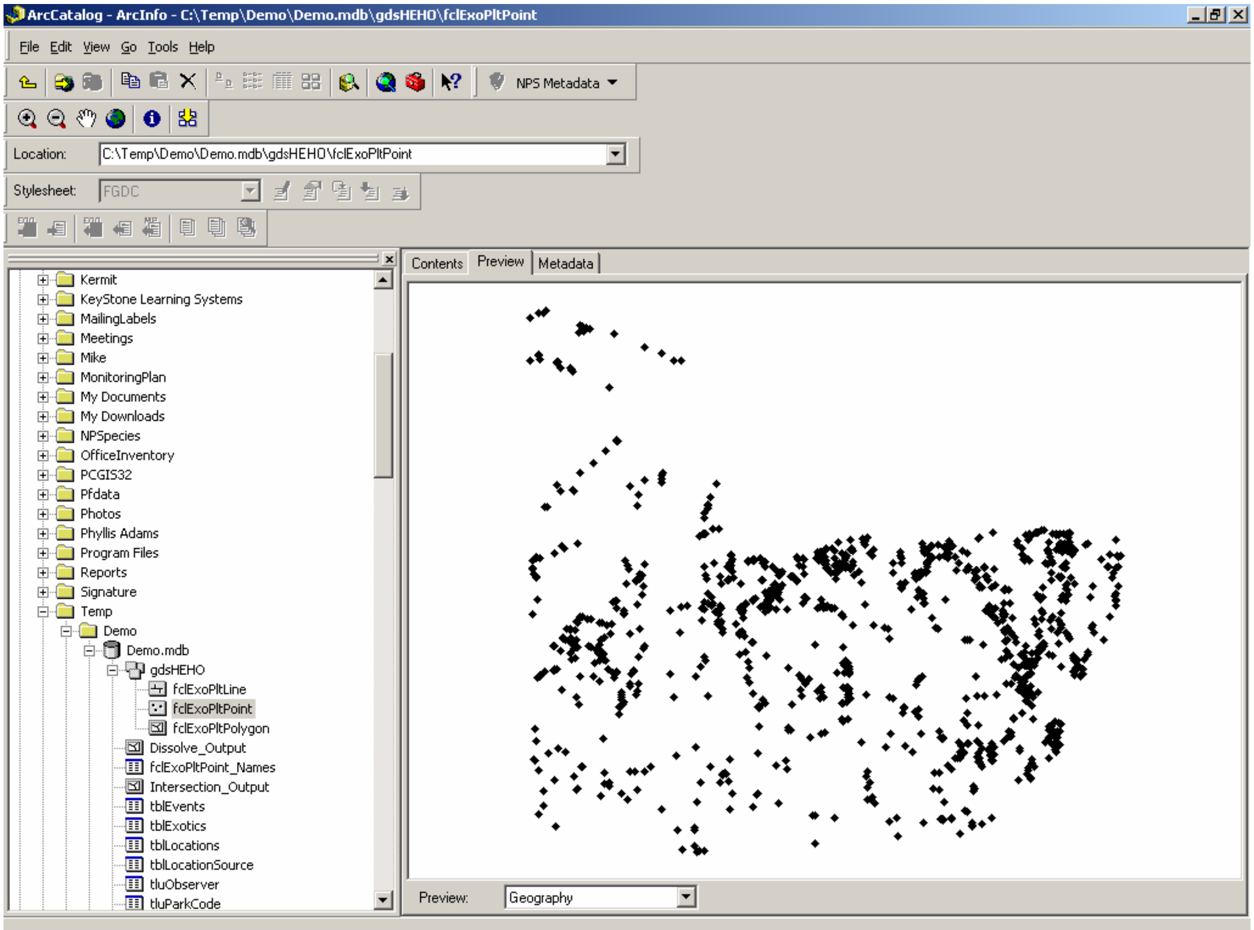


Figure 6. Typical ArcCatalog view of gdb components showing feature datasets, feature classes, and attribute tables.

5. Data Management Process

This section presents a conceptual model for data flow; then we describe a generalized process needed for the management of most Network I&M databases.

5.1 Data Flow Model

A simple data flow model provides a good starting place for understanding how data handling procedures are carried out at HTLN (Figure 7). The model identifies five key steps in data flow: acquisition, verification, validation, and analysis and dissemination. These steps plus additional procedural details are described in the following sections. Section 5.8 gives schedules for annual and long term data summary and analysis in monitoring projects. Storage, maintenance and security issues apply to all stages of the data flow.

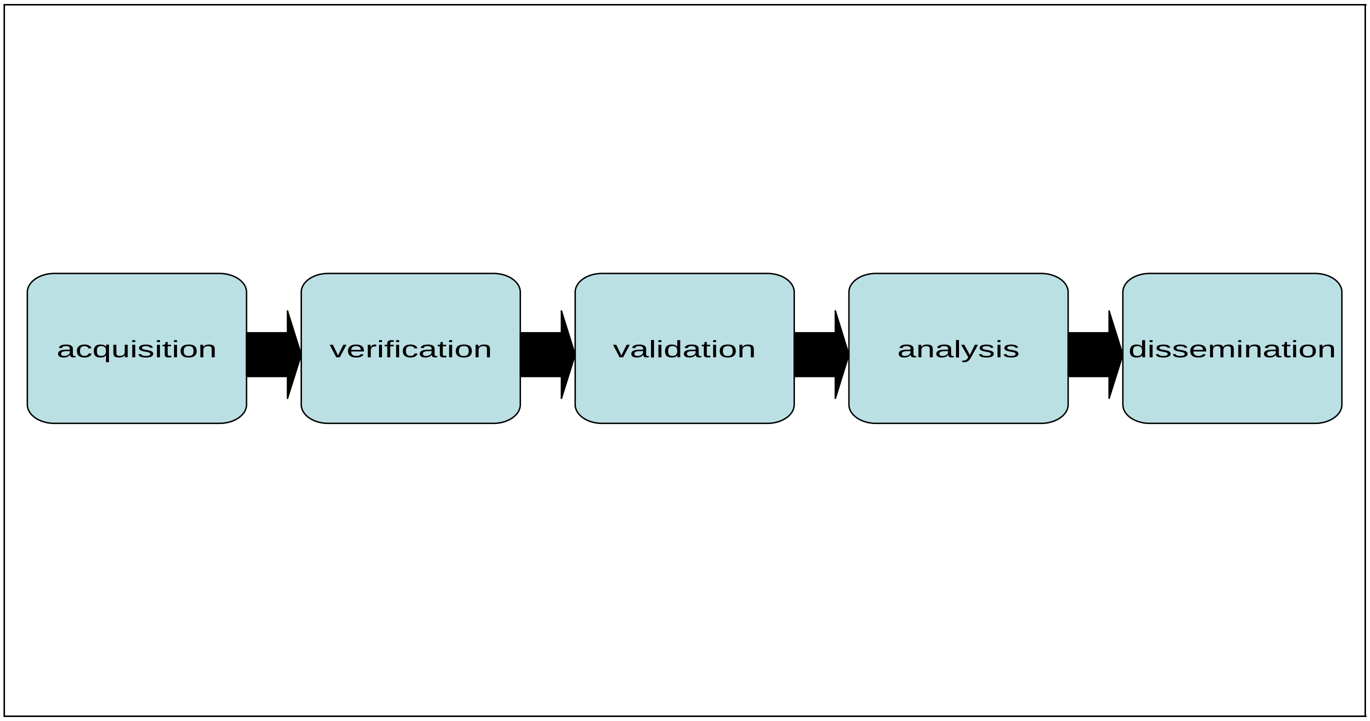


Figure 7. The flow of monitoring data can be viewed as a series of five basic management stages.

5.2 Data Acquisition

Data acquisition may also be the direct result of data mining or field studies conducted by Network staff. Development of the Network natural resource knowledge base involved an initial phase of data mining of existing documents and reports. Data mining included bibliographic lists of reports about monitoring conducted within and around each park. Additional data mining captured park related monitoring, especially that conducted by neighboring agencies. These reports were scanned and optical character recognition of legacy documents was completed and word-indexed through Adobe Capture and Acrobat. A database was then developed to summarize available information about monitoring activities within and around the Network parks.

Data acquisition of Network monitoring databases may include acquisitions through cooperative agreements with government agencies such as USGS-BRD or regional universities. Data acquisition may require data conversion from analog (usually paper) to digital or, on other occasions, be collected in digital form (GPS units, data loggers, sound recordings, photos). Cooperators should obtain NPS park permits prior to initiating monitoring studies, and both NPS staff and Cooperators should obtain state and federal fish and wildlife agency scientific permits prior to monitoring listed species. Spatially collected data follow accepted procedures (see GPS protocol SOP *in* HTLN Vital Signs Monitoring Plan: Phase III Suppl. Documents 5-10, 5-15 and 5-21) and, if required, be corrected using the nearest data correction station.

Data entry is the initial set of operations in which raw data from paper field forms or field notebooks are transcribed into a computerized form (i.e., within a database). Data entry commences as soon as data collection is complete. Data entry forms and QA/QC features minimize error. Data entry forms reduce transcription errors through pick lists and value limits and provide controlled access to the database. Forms also control the sequence of data entry. Inevitably, the process of transcribing data from field forms to a digital format introduces error. However, data entry forms and QA/QC features have been developed for each protocol to minimize error. For example, key fields are set to prevent duplicate entry of data.

Data entry forms reduce transcription errors through pick lists and value limits and provide controlled access to the database (i.e. forms are set for data entry only which prevents accidental deletion or alteration of existing data). Forms also control the sequence of data entry. For example, synonymous names (i.e. two or more different names referring to the same taxon) are common for plant species. Through the data entry form, a user searches for synonymous names before entering a new species name, thereby preventing the duplicate entry of synonyms. Figure 8 demonstrates some of the QA/QC features of the plant community data entry form.

5.3 Data Verification

Irrespective of how the data are acquired, the data are entered or imported into the database and a verification step is required. Data verification immediately follows data entry and involves checking the accuracy of digital data, primarily computerized records against the original source—usually paper field records. While the goal of data entry is to achieve 100% correct entries, this is rarely accomplished. To minimize transcription errors, our policy is to verify 100% of records to their original source by permanent staff. Further, 10% of records are reviewed a second time by the Project Manager and the results of that comparison reported with the data. If errors are found in the Project Manager's review, then the entire data set is verified again. Once the computerized data are verified as accurately reflecting the original field data, the paper forms are archived and the electronic version is used for all subsequent data activities

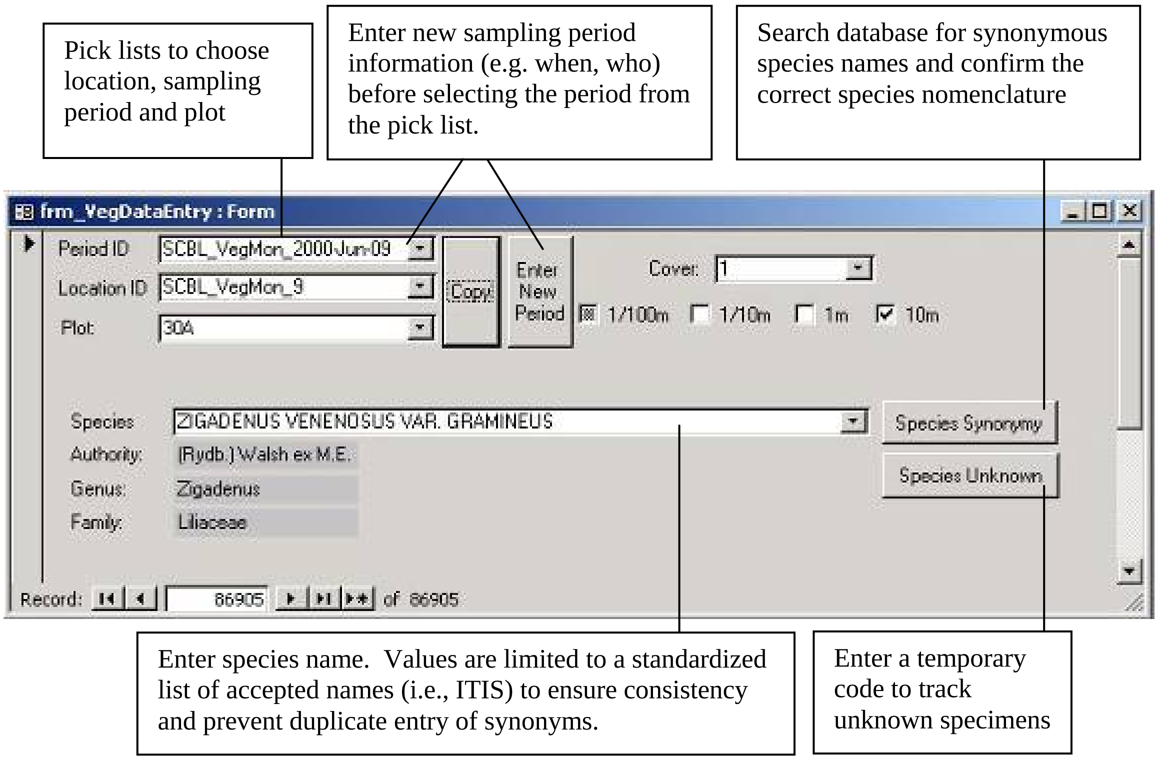


Figure 8. QA/QC features of the plant community data entry form.

5.4 Data Validation

Although data may be transcribed from the original field forms, they may not be accurate or logical. For example, a stream pH of 25.0 or a temperature of 95°C is illogical and almost certainly incorrect, whether or not it was properly transcribed from field forms. Also, spatial data that was collected within a park's boundary and when viewed in a GIS environment should be within the park's boundary. The process of reviewing computerized data for range and logic errors is the validation stage. Certain components of data validation are built into data entry forms (e.g. range limits). Additional data validation can be accomplished during verification, if the operator is sufficiently knowledgeable about the data. The Project Leader will validate the data after verification is complete. Validation procedures seek to identify generic errors (e.g. missing, mismatched or duplicate records) as well as errors specific to particular projects. For example, validation of plant community data includes database query and comparison of data among years. One query detects records with a location ID from a park and a period ID from a different park. Another query counts the number of plots per sample site (typically there are 10) to assure that all plots were entered. Finally, data are compared to previous years to identify gross differences. For example, *Dichanthelium oligosanthes* may be recorded this year, but *Dichanthelium spp* the previous. During the entry, verification and validation phases, the Project Leader is responsible for the data. The Project Leader must assure consistency between field forms and the database by noting how and why any changes were made to the data on the original field forms. In general, changes are made to field forms through marginal notes or attached explanations, not erasures. Once validation is complete, the data set is turned over to the Data Manager for archiving and storage.

Digital photos and sound recordings will be examined to determine their suitability for retention (i.e. only high quality photos will be kept) and evaluated against their metadata. Once photos are validated, each will be printed to high quality laserjet photo quality paper and documented appropriately (ANCS+, NPSpecies, photo database).

5.5 Data Organization

The various databases, reports, GIS coverages, etc. used and generated by the monitoring program create a large number of files and folders to manage. Several experiences from the Prairie Cluster reinforce the complicated nature of file management. For example, databases are occasionally stored in two versions of MS Access in order to accommodate data users with different software versions. Further, GIS data are sometimes stored in two projections – one for navigation, the other for use with existing base GIS data. Poor file organization can lead to confusion and data corruption. Figure 9 depicts an example file organization structure for the monitoring projects.

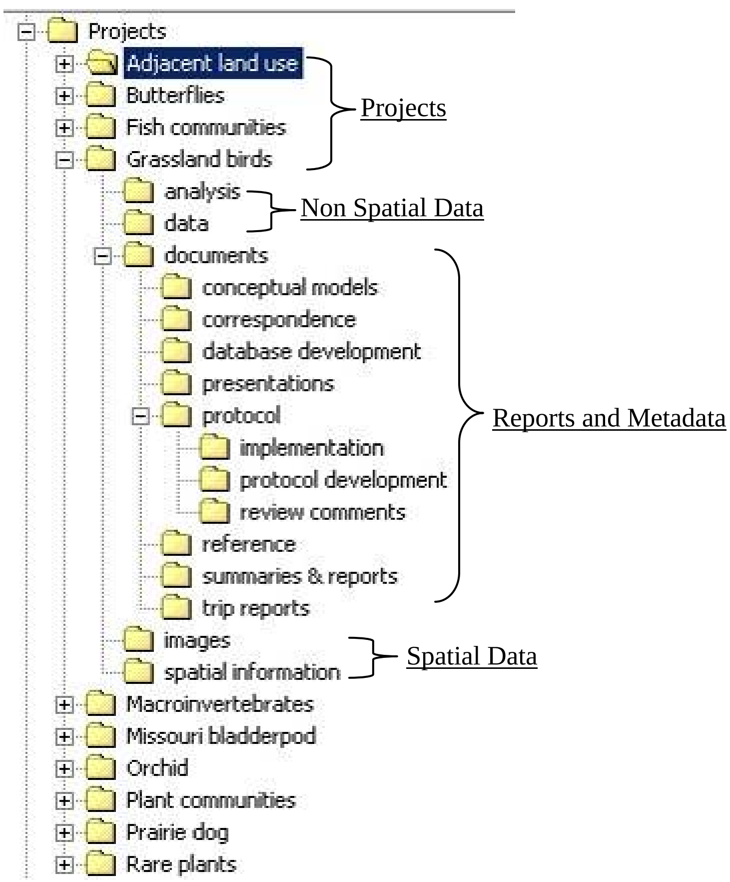


Figure 9. File structure for monitoring projects.

5.6 Data Maintenance and Storage

Once the data are archived, any changes made to the data must be documented in an edit log. At this point forward, original field forms are not altered. Field forms can be reconciled to the database through the use of the edit log.

Secure data archiving is essential for protecting data files from corruption. Once a data set has passed the QA/QC procedures specified in the protocol, a formal entry is made in the I&M Data Set Catalog. Analog data sheets are scanned using a high speed scanner and converted to Adobe .pdf documents. The analog data sheets are stored on-site. Subsequently, an electronic version of the data set is maintained in a read-only format on the program server. Backup copies of the data are maintained at a secure location (visitor's center), and an additional digital copy is forwarded to the NPS I&M archives.

Data Maintenance

Data sets are rarely static. They often change through additions, corrections, and improvements made following the archival of a data set. There are three main caveats to this process:

1) Only make changes that improve or update the data while maintaining data integrity.
2) Once archived, document any changes made to the data set.
3) Be prepared to recover from mistakes made during editing.

The Project Manager and Data Manager jointly edit archived data. Changes are documented in the edit log. Each change is accompanied by an explanation that includes pre- and post-edit data descriptions. The reader is referred to Tessler & Gregson (1997) for a complete description of prescribed data editing procedures and an example edit log.

Version Control

Prior to any major changes of a dataset, a copy is stored with the appropriate version number. This allows for the tracking of changes over time. With proper controls and communication, versioning ensures that only the most current version is used in any analysis. Versioning of archived data sets is handled by adding a three digit number to the file name, with the first version being numbered 001. Each additional version is assigned a sequentially higher number. Frequent users of the data are provided with a copy of the most recent archived version.

Data Security

Project managers maintain the current data files for their project. These data files are stored on the office server. The server uses a RAID 5 configuration to ensure data integrity. Access to these files is controlled through the use of user privileges. Typically, the Data Manager has universal access to data sets, while access for Project Managers is limited to their projects. Prior to storage on the server all files are scanned for viruses. Additional security is provided through tape backups. A complete server dump is completed at least two times per week. Three tape copies are stored locally and four copies are stored off-site. Tape copies are "rotated" between the Program Building and at a secure offsite location.

5.7 Data Analysis

The process of data analysis follows, but is not limited to, five basic steps: data determination, exploratory data analysis, the statistical test, a posteriori power analysis, and interpretation of the results. This will aid in planning the proper course of data analysis-simply put, there are a multitude of statistical tests but only a few will be appropriate. Example tools for analyses include Erdas Imagine, MS Access, ArcGIS, SAS, NCSS/PASS.

5.7.1 Overview

The approach towards data analysis and interpretation of results should be in accordance with scientific standards and appropriate methods and should be evaluated by way of statistical significance and power analyses.

5.7.2 Analysis for Annual Reports

Summary analysis for annual reports of Vital Signs monitoring studies should include descriptive statistics (mean, standard deviation, sample size) for all of the primary variables included in the project. Indices for calculating species richness, diversity and evenness for plant and animal community data can also be useful. Two indices, used by our network for animal and plant community data, are the Shannon diversity index, H' (Shannon, 1949; Magurran, 1988) and species distribution evenness, J' (Pielou, 1969).

An application of these indices is given in the Prairie Cluster Prototype Monitoring Protocol for Grassland Birds (Peitz et al., 2003). Bird diversity, richness and distribution evenness are calculated by plot with means (± SE) estimated by habitat from these calculations. Bird diversity for each plot is calculated using Shannon Diversity Index:

$H' = -\Sigma(n_1/N)\ln(n_1/N)$, were n_1/N is the proportion of the total number of individuals in a population consisting of the i^{th} species (Shannon, 1949). Species richness is determined as the total number of bird taxa recorded per plot. Species distribution evenness is calculated by plot using Pielou (J'):

$J' = H' / H_{max}$, where H' is the Shannon Diversity Index and H_{max} is the maximum possible diversity for a given number of species if all species are present in equal numbers (ln(species richness)). J' is a measure of how evenly individuals are distributed within a community when compared to the equal distribution and maximum diversity a community can have (Pielou, 1969).

The summary analyses of monitoring data should be complete, descriptive and easily interpretable. Measured variables, such as those characterizing the habitat or environment can be analyzed with respect to the primary variables using correlation matrices. Both the data analysis and annual report should provide resource managers timely feedback to help assess management practices.

To facilitate the timeliness of annual reporting, much of the summary analysis can be automated using queries and reports in MS Access. Measures that can be easily automated include frequency or abundance totals by species, species diversity, richness and distribution evenness by habitat, and descriptive statistics for associated variables. The summary reports can be exported in Word format and included as part of the project annual report.

The following example of summary automation is taken from the monitoring protocol: Fish Community Monitoring in Prairie Park Streams with Emphasis on Topeka Shiner (*Notropis topeka*) (Peitz and Rowell, 2003). Annual data summaries and reports for fish observation data are generated using the front-end within the database called “Shiner1.1”. When Shiner1.1 is opened, a switchboard form is displayed which includes the heading “National Park Service Inventory and Monitoring – Fish Community Monitoring: Fish Surveys”. Fish observation summaries are obtained by clicking on various Data Summaries and Reports buttons (see Figure 10). The data summary form is divided into fish community summary reports, Topeka shiner summary reports, and habitat summary reports. Habitat parameters are summarized for each stream reach and also for individual pools within each stream reach. Details about the interface design are given in Peitz and Rowell (2003) SOP#10 Data Analysis. The output of each option is an Access report that can be used for further statistical analysis or included directly into annual reports.

5.7.3 Long-term Analysis of Monitoring Data

The stated purpose of the Vital Signs monitoring program is to design and implement long-term ecological monitoring in order to provide information for park managers to evaluate the integrity of park ecosystems and better understand ecosystem processes. The approach towards long-term analysis of monitoring data is therefore critical in order to meet this goal.

Our network is currently planning and implementing several analysis techniques to address long-term data analysis for monitoring projects. As a working definition, we define long-term to be three or more years. Here, we present some types of long-term data analysis currently used in monitoring of grassland bird communities, prairie-stream fish communities, and vegetation monitoring in short/tallgrass prairie. Future directions for analysis are also discussed.

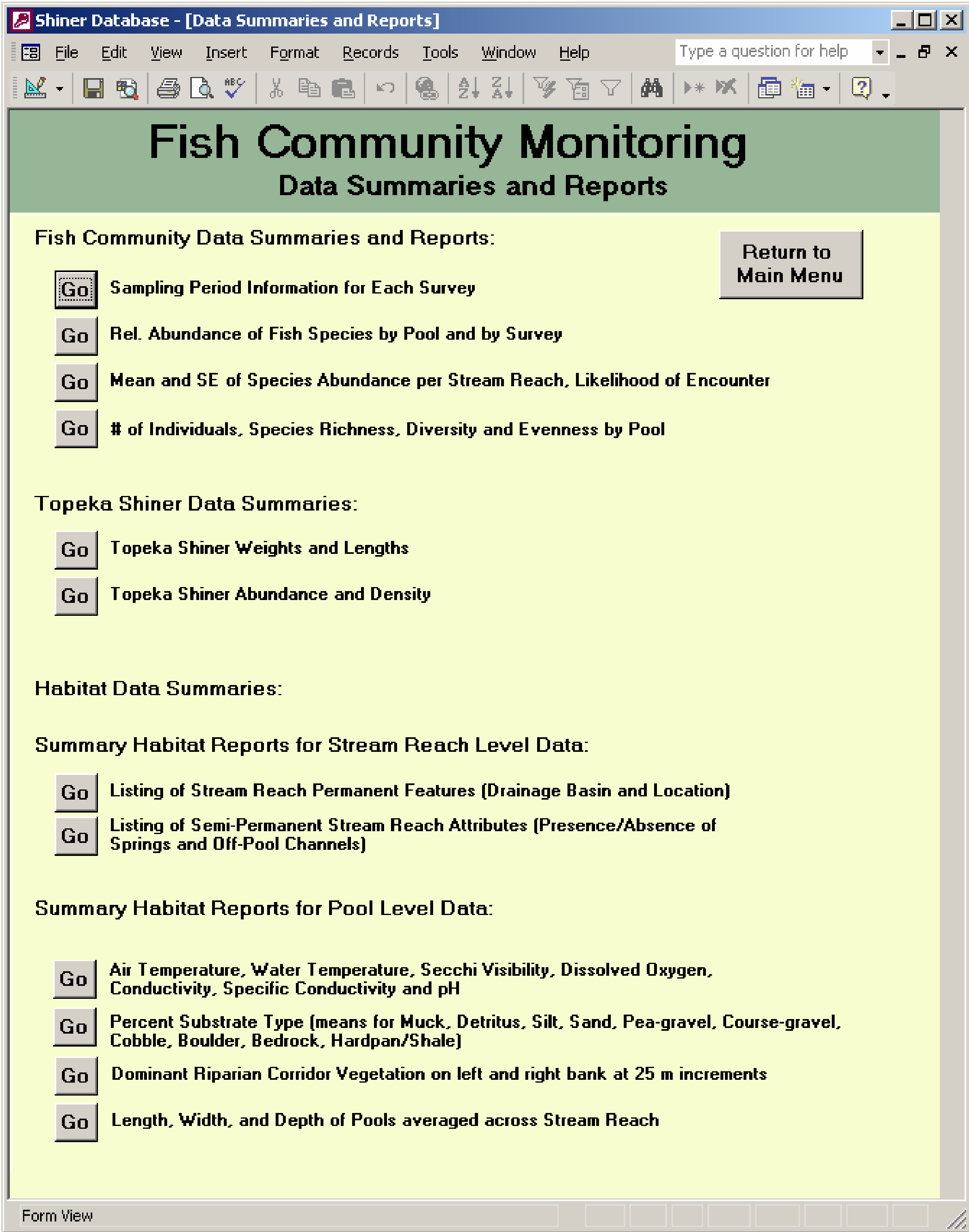


Figure 10. Data Summaries and Reports form for the Fish Community Monitoring database front-end. Options in the upper half of the form provide summaries of fish observation data. Options in the lower half of the form provide summaries of habitat data.

Measuring Diversity Trend in Prairie Plant Communities

The simplest, yet, perhaps, most important method of long-term monitoring data analysis is trends analysis. Trends in prairie plant communities addresses an array of ecological measures including species richness and diversity, exotic species relative frequency and cover, plant community composition including relative frequency and cover of plant guilds, plant community structure including ground cover and vegetation type cover, and plant community composition based on frequency of herbaceous and shrub species. Trends in plant community composition based on guilds, species diversity, species richness, and exotic to native species ratios are currently being explored.

Species-Habitat Correlation Analysis

For monitoring studies addressing grassland bird communities and prairie-stream fish communities, the relationship between species abundance and various habitat components have been monitored over several years. Avian species abundance may be correlated with a number of measurable habitat components depending on local circumstances. For grassland birds, we are tracking twelve biotic and abiotic habitat variables at bird observation sites.

Where sample sizes are sufficient, fish species abundance may be compared with habitat factors such as temperature, pH, oxygen, conductivity, and pool size and substrate. Significant associations between fish species abundance and habitat correlates will be identified where they exist. Partial correlation coefficient analysis may be used to examine habitat requirements for certain fish species where multiple physical factors appear to affect their abundance.

Other Data Analysis Techniques (in development)
Several vital signs will have monitoring implemented between FY04-08. These include metal contaminants, exotic species, and T&E species. Anticipated annual analyses include summary and descriptive statistics, and map development. Currently, there are three techniques being developed: data visualization, geostatistics, and network analysis.

Data visualization – We are currently considering the use of USGS Digital Elevation Models (DEMs) and Digital Ortho Quarter Quads (DOQQs) to visualize the relationship between prairie stream drainage physical features and the structure of fish communities that occur there. In particular, a surface viewshed of neighboring drainages may help to elucidate the relative roles of slope, aspect, and isolation (either by distance or by barriers such as dams on livestock ponds) in shaping the fish community structure.

Geostatistics – We have an ongoing project to measure the spatial variability of grassland birds and plant communities at Tallgrass Prairie National Preserve. The plan is to measure the spatial variability in species abundance, species diversity and associated environmental information as derived from IKONOS imagery. Geostatistical models will be combined with GIS analysis of differentiation diversity indices to measure species diversity change. The goal is to see how geostatistics and other spatial indices can contribute to our understanding of the spatial variability in park natural resources and the significance of spatial variability to natural resource management.

Network Analysis – Another GIS approach, network analysis, may assist in quantifying effective isolation by distance among sampling pools in water resource-related studies, such as the Network's prairie stream fish community project. USGS National Hydrographic Datasets (NHD) are available for the Tallgrass Prairie vicinity. NHD provides continuous network data for Cottonwood River watershed which includes the streams at Tallgrass Prairie. Hydrography data exists for Pipestone Creek, also part of the study, in the form of 1:24,000 USGS DLGs. Unlike the NHD datasets, unfortunately, DLGs do not provide continuous watershed linear features. Significant investment in data conversion would be required to prepare these data for network analysis. USGS plans to convert southwest Minnesota watersheds including the Pipestone Creek watershed into NHD data format in the near future.

Other quantitative methods may be of interest for analysis of long-term monitoring data including time-series analysis (Chatfield, 2004), "distance analysis" for animal population density estimates (Buckland et al. 2001), and Bayesian approaches for developing spatial and spatio-temporal models (Carlin and Louis, 2000).

5.8 Data Management Necessary for Analysis

5.8.1 Overview

Network staff are involved in multiple concurrent monitoring projects, so coordination is necessary to provide the required level of support for analysis and reporting. Our network will follow a schedule of project milestones and dates to meet the monitoring objectives (see Table 5). Reports for each Vital Sign will be written by Project Leaders in cooperation with Data Managers. Reports will include annual data summaries written for Park and Network Managers, five-year trend reports for Park Superintendents and Washington and Regional Natural Resource Managers, internet websites for NPS staff and the general public, and e-mail bulletins for Park Superintendents, Natural Resource Managers, and the general public on-request.

5.8.2 Identifying Project Objectives and Data Products

Identifying project objectives for existing and planned vital signs monitoring projects is usually the first step in developing data products. Monitoring databases are almost always the outcome of field studies associated with the vital signs projects. The plan is that databases will generally come directly from Project Leaders and be as complete as possible. Still, additional data will often be required in order to complete the necessary analyses. Weather data, complete site information, taxonomic information, complete names of observers, and even database design changes such as keys, indexes, and constraints may be needed prior to analysis. The development of data products will be guided by the Vital Sign project objectives, and when available, by the protocols and data management and data analysis SOPs (see HTLN Vital Signs Monitoring Plan: Phase III Suppl. Documents 5-10, 5-15 and 5-21 for examples).

Table 5 – Data Analysis and Reporting Milestones for HTLN Vital Signs Monitoring Projects.

Level 1	Level 2	Level 3	Network Vital Sign Name	Data Collection	Data Entry, Verification and Validation	Data Analysis / Reporting	Data Summary Automated	Data Archival	Long-term Trends Reporting
Air and Climate	Air Quality	Ozone	Ozone	Continuous	Continuous	Continuous	Yes	Continuous	10 years
		Wet and Dry Deposition	Wet and Dry Deposition	Continuous	Continuous	Continuous	Yes	Continuous	10 years
		Visibility and Particulate Matter	Visibility and Particulate Matter	Continuous	Continuous	Continuous	Yes	Continuous	10 years
		Air Contaminants	Air Contaminants	Continuous	Continuous	Continuous	Yes	Continuous	10 years
	Weather and Climate	Weather and Climate	Weather	Continuous	Continuous	Continuous	Yes	Continuous	10 years
Geology And Soils	Geomorphology	Stream/ River Channel Characteristics	Fluvial Geomorphology	TBD*	TBD*	TBD*	TBD*	TBD*	TBD*
			Stream Habitat/ Riparian Assessment	1 June – 1 Oct	1 Mar	15 Apr	Yes	30 Apr	10 years
		Surface Water Dynamics	Stream Discharge	Continuous	Continuous	Continuous	Yes	Continuous	10 years
	Water Quality	Water Chemistry	Core Water Quality Parameters	TBD*	TBD*	TBD*	TBD*	TBD*	TBD*
			Pollutant Metals	TBD*	TBD*	TBD*	TBD*	TBD*	TBD*
		Aquatic Macroinvertebrates and Algae	Aquatic Macronivertebrates—Prairie Streams	1 June – 1 Oct	1 Mar	15 Apr	Yes	30 Apr	5 years
			Aquatic Macronivertebrates—Rivers	1 Oct. – 1 Mar.	TBD*	TBD*	TBD*	TBD*	5 years

Level 1	Level 2	Level 3	Network Vital Sign Name	Data Collection	Data Entry, Verification and Validation	Data Analysis / Reporting	Data Summary Automated	Data Archival	Long-term Trends Reporting
Biological Integrity	Invasive Species	Invasive/Exotic Plants	Exotic Forest Plants	15 Apr – 15 Oct	TBD*	TBD*	TBD*	TBD*	TBD*
		Invasive/Exotic Plants	Exotic Grassland Plants	15 Apr – 15 Oct	TBD*	TBD*	TBD*	TBD*	TBD*
	Focal Species or Communities	Wetland Communities	Wetland Plant Communities	TBD*	TBD*	TBD*	TBD*	TBD*	TBD*
		Grassland Communities	Prairie Community Structure, Composition, and Diversity	15 Apr – 15 Oct	15 Jan	1 Mar	Yes	1 Apr	5 years
		Forest Vegetation	Forest Community Structure, Composition, and Diversity	15 Apr – 15 Oct	15 Jan	1 Mar	Yes	1 Apr	5 years
		Fishes	Fish Community—Prairie Streams	1 Sep – 15 Oct	31 Dec	1 Mar	Yes	15 Mar	3 – 5 years
			Fish Community—Ozark Rivers	1 Oct. – 1 Mar.	TBD*	TBD*	TBD*	TBD*	5 years
		Birds	Landbirds	1 May – 30 Jun	31 Dec	1 Mar	Yes	15 Mar	3 – 5 years
		Mammals	Deer	TBD*	TBD*	TBD*	TBD*	TBD*	TBD*
			Missouri Bladderpod	1 Sep – 15 Jun	31 Dec	1 Mar	Yes	15 Mar	5 years
			Ozark Hellbender	TBD*	TBD*	TBD*	TBD*	TBD*	TBD*
			Topeka Shiner	1 Sep – 15 Oct	31 Dec	1 Mar	Yes	15 Mar	3 – 5 years
			Western Prairie Fringed Orchid	1 Jul – 31 Jul	31 Dec	1 Mar	No	15 Mar	5 years
		Land Cover and Use	Land Cover/Land Use	NA	NA	NA	NA	NA	10 years

* To be determine through ongoing protocol development efforts.

5.8.3 Data extraction

Project scopes-of-work, and whenever possible, annual reports from previous years will be used to identify the tables and fields necessary for data summary. Assuming a database already exists, one can begin by identifying the core or critical tables with the desired data addressing the project goals. Then one may review the total database design including Entity-Relationship Diagrams (ERDs), priority fields, and the tables that contain them, to work towards building queries for data summary and analysis. These are also the preliminary steps that will be used for report automation (below).

Individual fields will be identified which provide the source data for the desired analyses. Very often, the fields are to come from two or more tables. So it will become necessary to join tables using queries. ERDs will be used to determine the most effective way to write queries. Usually analysis queries do not yet exist, and these will be built by the Data Manager in coordination with the Project Leader. The Project Leader will start by identifying one or two variables to be used in the analysis, and then add independent variables that will be used for grouping purposes such as site, habitat, or time interval. These will form the basis for the analysis queries. The analysis will be developed through a series of queries until either the data are exported to outside applications (see below) or the desired result is calculated.

5.8.4 Automated Reporting and Data Summaries

The automation of data summaries and annual reports facilitates our network's ability to manage multiple projects. The Network uses Microsoft Access to automate its data summaries and reports although the process would be basically the same in other application environments.

The development of automated reporting would be greatly facilitated by existing annual reports when these are available. If not, a detailed annual report template will be worked out between the Project Leader and the Data Manager. The job of the Data Manager will be to take to the extracted tables and fields and to process them to create output that is similar, if not identical, to the existing annual report template. This task requires that the Project Leader have a very clear idea of what will be needed for the annual reports. Usually, the Data Manager will need only to develop the necessary queries, macros, modules and reports. The process of automation is then to link the various database objects to a simple user-interface. Examples are given in the Data Analysis SOPs in HTLN Vital Signs Monitoring Plan: Phase III Suppl. Documents 5-10, 5-15 and 5-21.

Often, the Project Leader will need to modify the existing annual reports to better meet the requirements of Vital Signs projects. The Data Manager will plan on expanding and improving the analysis and reporting products during the automation phase, if possible, because the underlying database objects will have to be thoroughly reviewed anyway. Developing the automated reporting system beyond the existing annual reports will require frequent and repeated interviews by the Data Manager with the Project Leader. Once again, the project objectives will facilitate communication between the Data Manager and the Project Leader. The Network will often sketch out hardcopies of the data summaries, annual reports and user-interfaces in pencil prior to starting the implementation phase.

Following discussions on data summary and reporting requirements, the next step is to implement the automated report. Except in the simplest of situations, we will develop each analysis independently through a series of Access queries. The resultant query will be loaded as one of several controls (textboxes) into the summary report. All queries that feed into a single report may be triggered from a form button, a macro, or a combination of both. The Network will use modules and VBA for automation only to the extent that more complex automation applications are required.

5.8.5 Exports to statistical packages and other software

At times, data will need to be exported out of the database to other software applications. The Network is planning to export data from Access databases for most statistical analysis beyond means, standard deviations, and other descriptive statistics. The Network will be using third party statistical software for frequency distribution plots, tests for normality and analysis of variance such as SAS, SPSS and NCSS. Other external software requiring data exports will most likely include special application software such as GS+ for geostatistical analysis and Distance 4.0 for estimating animal population abundance using distance sampling.

Analysis objectives will be reviewed by the Data Manager and Project Leader together prior to attempting data exports. The field order [order of variables in the resulting ‘flat’ file] is most easily controlled while the data are still in the relational database. Therefore, the Data Manager will assist in setting up the necessary queries to provide the field order. Field types are typically determined early on, when the database is being designed. To minimize risk of data loss as a result of data type conversion, changes in data types will be kept to a minimum once monitoring projects are underway.

ASCII text has the advantage of being almost universally readable by third party applications. Typically, fields are delimited by commas, tabs, or spaces. Text strings are usually enclosed by single or double quotes. ASCII text requires extra steps transferring data between applications and also extra care regarding data format. The Network will be using ASCII format only when other options are not available. The preferred alternatives are data exports directly from an Access .mdb file (such as ESRI ArcGIS) or link to an .mdb file by way of a data base connection such as OLE DB or ODBC data link (many Windows-based applications).

5.8.6 Data flow for subsequent analysis by third parties

In order to create a data flow for subsequent analysis by third parties, the Network will establish a timeline and data processing steps including error-checking, summarizing, analyzing, and distributing data. A data flow model including error-checking, analyzing, reporting, and distribution is described in detail earlier in this document (Sections 5.1 – 5.6) Timelines and report dissemination issues are described below. Figure 8.1 summarizes an approach where each project leader is responsible for their dataset during the course of the year, but once a year they review the dataset, clean it up, do their annual report, and make the data available in a common repository for others to do synthesis and further analysis. Specific project examples follow that describe when each project leader or the data manager complete these tasks to make their particular data set available to others each year.

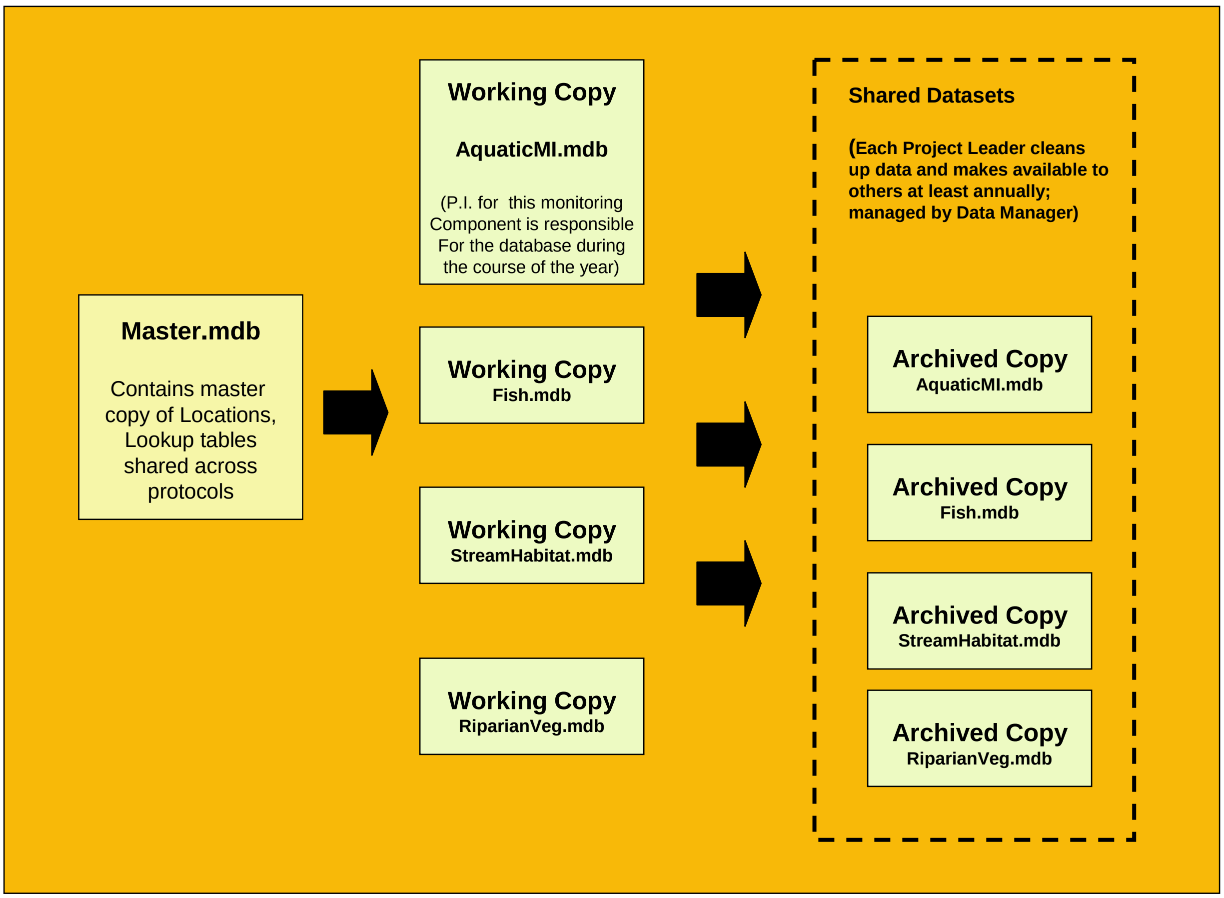


Figure 11. Data flow from program databases to shared datasets; to be made available to others for further analysis and synthesis (based on S. Fancy, pers. comm.).

5.8.7 Specific examples

Aquatic Macroinvertebrates

Benthic macroinvertebrate monitoring is being conducted at Wilson's Creek NB to determine annual status of stream macroinvertebrate communities in order to assess the overall biotic integrity of Wilson's Creek and to detect changes through time in macroinvertebrate communities. Samples are identified and enumerated and taxa richness, family richness, family diversity, EPT richness, EPT ratio, and family biotic index are determined. Additionally, effluent discharge and chemistry (incl. total phosphorus, and total nitrogen) is also monitored (by a non-GOV agency) and monthly total phosphorous data from Wilson's Creek are analyzed. Data collection takes place from 22 May through 19 September each year. Data entry and error checking are completed by March 1st. Reports and archiving of shared datasets are completed by April 15th .

Grassland Birds

Grassland birds were initially selected for monitoring as indicators of overall prairie ecosystem health. Grassland bird inventories were conducted in the parks in 1998 and 1999 as a preliminary step toward developing long-term monitoring. We have implemented bird community monitoring at Agate Fossil Beds NM, and Tallgrass Prairie NP, the only parks where grassland bird species represented a relatively high proportion of bird species present. Questions we are addressing include: What is the current status in grassland bird populations? What are the long-term abundance trends? What is the current condition of prairie habitat, and how is it changing through time? Data collection takes place from 1st of May through 30th of June each year. Data entry and error checking are completed by December 31st. Reports and archiving of shared datasets are completed by March 1st.

Grassland Plant Communities

In many of the Network parks, grassland plant communities are important natural resources and the focus of much management attention. Although small in size, our parks represent the few remaining refuges where the once widespread prairie grasslands persist. Furthermore, intact prairie represents the historical landscape context for the cultural resources the parks are intended to interpret. Profound alteration of lands in the mid-continent to agricultural use has permanently disrupted the natural forces of wildfire and grazing. Consequently, managers employ prescribed fire, manual removal of woody species, exotic control and restoration to maintain the prairie. To date, the effectiveness of management actions in sustaining prairie in the face of fragmentation, disruption of natural disturbance regimes, and exotic species encroachment is uncertain. The purpose of monitoring is to track the current species composition, structure, and diversity of remnant and restored prairies. What are trends in species composition, structure, and diversity and how are they correlated with climatic variables or management activities, such as prescribed fire? Data collection takes place from 15th of April through 15th of October each year. Data entry and error checking are completed by January 15th. Reports and archiving of shared datasets are completed by March 1st.

Threatened and Endangered Species

Three federally listed threatened and endangered species and one former candidate species are currently monitored by the Network. These include western prairie fringed orchid (listed – threatened), Topeka shiner (listed – endangered), Missouri bladderpod (listed – threatened), and black-tailed prairie dog (removed from candidate list, Federal Register, July 19, 2004).

Western Prairie Fringed Orchid (*Platanthera praeclara*)
Population size, density, and composition and responses to management regimes are being monitored for the western prairie fringed orchid at Pipestone NM. The analyses being conducted provide a record for any changes in the abundance and distribution of flowering individuals through time. A secondary objective of the monitoring is to estimate the reproductive capability of the population. Number of plants, plant height, and number of flowers/floral buds are collected and annual means and standard deviation are calculated for plant height and number of flowers/floral buds. Plant locations are recorded to determine changes in distribution. Analysis of variance (ANOVA) and a means comparison test is used to determine significant differences in mean inflorescence height and mean number of flowers/buds per inflorescence among census years. Trend results are depicted via graphics and distribution maps are included with the report. Data collection takes place from 1st of July through 31st of July each year. Data entry and error checking are completed by December 31st. Reports and archiving of shared datasets are completed by March 1st.

Topeka shiner (*Notropis topeka*)
Species richness and relative abundance of fish communities, as well as three Topeka shiner parameters -relative abundance and distribution, reproductive success, habitat quality and quantity - are being monitored at Pipestone NM and Tallgrass Prairie N Preserve. Number of fish, by species, is collected and summarized by reach, Topeka shiners are measured, sexed and aged, and a map is produced depicting monitoring sites. Additionally, air and water temperature, water clarity, dissolved oxygen content, conductivity and pH, substrate cover by types, presence and type of stream bank vegetation, stream bank erosion level, substrate condition, and pool dimensions data is collected and will be summarized. Analyses include means, ranges, and standard deviations. The analyses provide a means to monitor through time the status and trends in each of the above parameters. Data collection occurs from 1st of May through 15th of October each year. Data entry and error checking are completed by December 31st. Reports and archiving of shared datasets are completed by March 1st.

Missouri Bladderpod (*Lesquerella filiformis*)
Population size, density, and composition (spatial and biological) and responses to management regimes are being monitored for the Missouri bladderpod at Wilson's Creek NB. The analyses being conducted estimate bladderpod abundance and fluctuations over time, bladderpod occurrence, survivorship and reproduction as they vary with habitat characteristics, and changes in limestone glade habitat over time. Data collection takes place from 1st of September through

15th of June each year. Data entry and error checking are completed by December 31st. Reports and archiving of shared datasets are completed by March 1st.

Black-tailed Praire Dog (*Cynomis ludovicianus*)
Prairie dog density, abundance and colony size are monitored at Scotts Bluff NM. Two calculations are made to estimate annual density and abundance based on visual counts. A 95 % confidence interval is also calculated for density and abundance. Additionally, annual colony size and location are mapped and Sylvatic plague (*Yersinia pestis*) is evaluated in prairie dog colonies. Anticipated trend analyses are still being developed. Data collection takes place from 1st of May through 15th of August each year. Data entry and error checking are completed by December 31st. Reports and archiving of shared datasets are completed by March 1st.

5.9 Dissemination

Following the appropriate review process, dissemination of data, reports, and other items (photos, sound recordings, etc.) will be in accordance with current NPS standards. Species lists and voucher data will be input to NPSpecies and ANCS+ and subsequently updated (if required) whereas reports will be documented in NatureBib. Natural resource data and GIS data will be forwarded to the Natural Resource and GIS Data Store, a web-based application designed to integrate data dissemination and metadata maintenance for Natural Resource and GIS datasets, digital documents and appropriate digital photos. Reports are currently being converted to Adobe .pdf files and posted on the Network I&M web page http://www1.nature.nps.gov/im/units/htln/index.htm or at the Biodiversity Data Store http://science.nature.nps.gov/im/inventory/biology/BiodiversityDataStore.htm. Requests for data and additional information may be made by managers, researchers, and the public by calling or e-mailing staff listed on the contacts web page.

Monitoring geodatabases will be forwarded to the monitoring protocol website for distribution following QA/QC and metadata review (see above). In addition, completion and review of monitoring protocols and standard operating procedures (SOPs) will, most likely, be required prior to data collection. Special requirements for entering and managing water quality data in Storet follows the USGS-WRD and EPA guidance, http://www.epa.gov/storet/index.html.

5.10 Data Ownership and FOIA

(modified from NCBN Data Management Plan Chapter 9, Draft 2004).

All data collected are public property and subject to requests under the Freedom of Information Act (FOIA). The NPS policy regarding FOIA is discussed in Director's Order 66 (which is in draft form at this time). Exemptions from FOIA are discussed in detail in the Director's Order. Of specific concern to data managers of NPS inventory and monitoring data are: cave locations and resources, federally listed endangered or threatened animals and plants, and state listed endangered or threatened animals and plants. Certain cultural and water resource data are also exempt from FOIA so data managers need to become familiar with Director's Order 66 prior to distributing information. Any type of data dissemination will undergo an internal evaluation of their sensitivity status. Data that have no cave locations and resources, listed taxa, or other restrictions should be publicly available. Data that do have these restrictions will be evaluated with park specific staff to determine if the sensitivity will be park only or NPS only.

5.10.1 National Park Service Policy on Data Ownership

The National Park Service defines conditions for the ownership and sharing of collections, data, and results based on research funded by the United States government. All cooperative and interagency agreements, as well as contracts, should include clear provisions for data ownership and sharing as defined by the National Park Service:

- All data and materials collected or generated using National Park Service personnel and funds become the property of the National Park.

- Any important findings from research and educational activities should be promptly submitted for publication. Authorship must accurately reflect the contributions of those involved.

- Investigators must share collections, data, results, and supporting materials with other researchers whenever possible. In exceptional cases, where collections or data are sensitive or fragile, access may be limited.

The Office of Management and Budget (OMB) ensures that grants and cooperative agreements are managed properly. Federal funding must be disbursed in accordance with applicable laws and regulations. OMB circulars establish some degree of standardization government-wide to achieve consistency and uniformity in the development and administration of grants and cooperative agreements. Specifically, OMB Circular A-110 establishes property standards within cooperative agreements with higher institutions and non-profit organizations. Section 36 of Circular A-110, "Intangible Property" describes the following administrative requirements pertinent to data and ownership.

5.10.2 Establishing Data Ownership Guidelines

Our Network has established guidelines for the ownership of data and other research information. To ensure that proper ownership, format, and development of network products is maintained, all cooperative or interagency work must be conducted as part of a signed collaborative agreement. Every cooperative or interagency agreement or contract involving the Network must include OMB Circular A-110 cited under the *Reports and Deliverables* Section of all agreements and contracts. The following shows appropriate language to use when citing Circular A-110:

"As the performing organization of this agreement, <u>*institution or organization name*</u> shall follow the procedures and policies set forth in OMB Circular A-110."

Every cooperative or interagency agreement or contract must include a list of deliverables and products clearly defined within each agreement or contract. Details on formatting and media types that will be required for final submission must be included. Agreements and contracts must list all products expected to result from the project. These include, but are not limited to, field notebooks, photographs (hardcopy and digital), specimens, raw data, and reports.

5.11 Data Distribution
(modified from NCBN Data Management Plan, Chapter 9, Draft 2004).

One of the most important goals of the Inventory and Monitoring Program is to *integrate natural resource inventory and monitoring information into National Park Service planning, management, and decision making.*

To accomplish this goal, procedures must be developed to ensure that relevant natural resource data collected by NPS staff, cooperators, researchers and the public are entered, quality-checked, analyzed, documented, cataloged, archived, and made available for management decision-making, research, and education. Providing well-documented data in a timely manner to park managers is especially important to the success of the Program. The Network will make certain that:

- Data are easily discoverable and obtainable
- Data that have not yet been subjected to full quality control will not be released
- Distributed data are accompanied by complete metadata that clearly establishes the data as a product of the NPS I&M Program
- Sensitive data are identified and protected from unauthorized access and inappropriate use
- A complete record of data distribution/dissemination is maintained

To accomplish this, a number of distribution methods will be used that will allow information collected and developed as part of the program become widely available to park staff and the public.

5.11.1 Data Classification: protected vs. public

All data and associated information from I&M activities must be assessed to determine their sensitivity. This includes reports, metadata, raw and manipulated spatial and non-spatial data, maps, etc... Network staff must carefully identify and manage any information that is considered sensitive. The Network must clearly identify and define those data needing access restrictions and those to make public.

The Freedom of Information Act, 5 U.S.C. § 552, referred to as FOIA, stipulates that the United States Government, including the National Park Service, must provide access to data and information of interest to the public. FOIA, as amended in 1996 to provide guidance for electronic information distribution, applies to records that are owned or controlled by a federal agency, regardless of whether or not the federal government created the records. FOIA is intended to establish a right for any person to access federal agency records that are not protected from disclosure by exemptions. Under the terms of FOIA, agencies must make non-protected records available for inspection and copying in public reading rooms and/or the Internet. Other records however, are provided in response to specific requests through a specified process. The Department of the Interior's revised FOIA regulations and the Department's Freedom of Information Act Handbook can be accessed at http://www.doi.gov/foia/ for further information.

In some cases, public access to data can be restricted. Under one Executive Order, Director's Order #66B, and four resource confidentiality laws, the National Parks Omnibus Management Act (16 U.S.C. 5937), the National Historic Preservation Act (16 U.S.C. 470w-3), the Federal Cave Resources Protection Act (16 U.S.C. 4304) and the Archaeological Resources Protection Act (16 U.S.C. 470hh), the National Park Service is directed to protect information about the nature and location of sensitive park resources. Through these regulations, information that could result in harm to natural resources, including endangered or threatened species, can be classified as 'protected' or 'sensitive' and maybe withheld from public release.

Classification of sensitive I&M data will be the responsibility of the Network staff, the park superintendents, and investigators working on individual projects. Network staff will classify sensitive data on a case by case, project by project, basis. They will work closely with investigators for each project to ensure that potentially sensitive park resources are identified, and that information about these resources is tracked throughout the project.

The Network staff is also responsible for identifying all potentially sensitive resources to principal investigator(s) working on each project. The investigators, whether Network staff or partners, will develop procedures to flag all potentially sensitive resources in any products that come from the project, including documents, maps, databases, and metadata. When submitting any products or results, investigators should specifically identify all records and other references to potentially sensitive resources. Note that partners should not release any information in a public forum before consulting with Network staff to ensure that the information is not classified as sensitive or protected.

Note that information already in the public domain can, in general, be released to the public domain.

5.11.2 Access Restrictions on Sensitive Data

Network staff are responsible for managing access to sensitive data handled by the Program. All potentially sensitive park resources will be identified and investigators informed working on Network projects that:

- All data and associated information must be made available for review by Network staff prior to public release in any format

- Any information classified as protected should not be released in any format except as coordinated with the National Park Service

The Network Coordinator or Data Manager identifies all potentially sensitive park resources to the principal investigator for each project. Reciprocally, the principal investigators for each project must identify any known references to potentially sensitive park resources.

For each project, the Network Data Manager provides a complete list of all references to potentially sensitive park resources in each park to the park superintendent for review. Each superintendent then determines which information should be protected.

When preparing or uploading information into any Network database, the Network staff ensures that all protected information is properly identified and marked. The Network staff work together to ensure that all references to protected information are removed or obscured in any reports, publications, maps, or other public forum.

Network staff will remove any sensitive information from public versions of documents or other media. They will isolate sensitive from non-sensitive data and determine the appropriate measures for withholding sensitive data. The main distribution applications and repositories developed by the I&M Program, NatureBib, NPSpecies, the NR/GIS Data Store and the Biodiversity Data Store, are maintained on both secure and public servers, and all records that are marked 'sensitive' during uploading will only become available on the secure servers. Assigning a sensitivity level to data when uploading to both the NPSpecies and NatureBib databases are discussed between the Parks and Network staff.

Thus, access to data on sensitive park resources can be limited to Network staff or research partners, and limits to how this data are subsequently released must also be clearly defined. It is crucial that the Network staff institute quality control and quality assurance measures to ensure that the person doing the uploading of records into the online applications is familiar with the procedures for identifying and entering protected information.

5.11.3 Public Access to Network Inventory and Monitoring Data

According to FOIA (specifically the 1996 amendments), all information routinely requested must be made available to the public via reading rooms and/or the internet. Network project data will be available to the public at one or more internet locations:

- The Network web site
- Public servers for the NPSpecies and NatureBib databases
- Public server for the Biodiversity Data Store
- Public server for the NR/GIS Data Store

The Network will regularly provide information about inventories and monitoring projects, including annual reports and detailed project reports through the Network web site. Information on species in the National Parks, including all records generated through the Inventory and Monitoring Program, will be maintained and assessable through the NPSpecies database. Bibliographic references that refer to National Park System natural resources will be assessable through the NatureBib database. Documents, maps, and data sets containing resource information from all sources, and their associated metadata, will be assessable through the

Biodiversity Data Store and/or NR/GIS Data Store. Each of these databases/repositories will be accessible via both a secure server and a public server, and the public can access all information in these databases except those records marked as 'sensitive.'

5.11.4 Data Availability

Both raw and manipulated data resulting from the Network's inventory and monitoring projects will be fully documented with FGDC compliant metadata and made available to the public via the Network's website. The metadata for all datasets will be made accessible to the public as soon as they are provided and verified by the Investigator(s) or project managers.

Datasets for short-term studies (inventories) will be provided to the public on the Network website, two years following the year the data were collected or following publication of the Investigator's results (whichever comes first). Long-term (monitoring) studies will be provided to the public in four-five year intervals, or when trends analyses have been completed and reported on by the network. (This will be specific to each network monitoring protocol, refer to the Network's Phase 3 plan for further information). Before data are posted, the Investigator or project manager will be asked to verify the final dataset and metadata if necessary. Once the Network staff and Investigator agree, the data will be made accessible to the public, provided no sensitive issues identified.

Network staff will notify investigators when datasets will be made available to the public. This will allow each investigator the opportunity to request in writing to further restrict access to the dataset by the public. Network staff will review the Investigator's request and determine whether the request will be granted and for how long the dataset will remain restricted.

5.11.5 Data Acquisition Policy

The Network will develop a dataset acquisition policy that will be made available to all Network website users who wish to acquire program data and information. This policy will include such things as:

- A mandatory web-based questionnaire to be completed before acquiring data (This questionnaire will allow Network staff the ability to maintain a distribution log specifying recipient name and contact information, intended use of data, export file format, delivery date and method, and data content description noting range by date and geography of data delivered.)

- A statement about use and appropriate citation of data in resulting publications

- Request that acknowledgement be given to the National Park Service Inventory and Monitoring Program within all resulting reports and publications

All data sets with public access available on the Network website will be accompanied by the Network's acquisition policy.

6. Database integration – Role of the Vital Signs Framework

Based on the data resources description in this document, the Network may be managing as many as 90 I&M databases (both inventory and monitoring data). It seems clear that the total number of databases could easily become unmanageable, and the original purpose of the program, to provide park managers with the scientific data they need to understand and manage park resources, could be missed altogether. The WASO office has been developing the Vital Signs Framework in order to meet this goal and bring a unifying structure to the I&M program (Fancy 2004). The Framework also serves to integrate the many databases in order that questions may be meaningfully asked of the vital signs monitored.

The Vital Signs Framework is a hierarchical classification of the monitoring vital signs. There are three levels in the Framework classification. An application of the Framework with Network I&M projects is given in Table 3. The classification allows data users to drill down across multiple databases by theme, with increasing specificity. For example, if several park managers are concerned about the state of endangered species among National Parks in a region, they can begin by selecting the Level 1 category Biological Integrity, at Level 2, select Focal Species or Communities, and at Level 3, T&E species and communities. This selection is cross-referenced by a list of parks. The resulting selection would be all of the data on T&E species at the parks of concern. The results could be further limited by including a range of dates if Table 3 is joined to specific monitoring data.

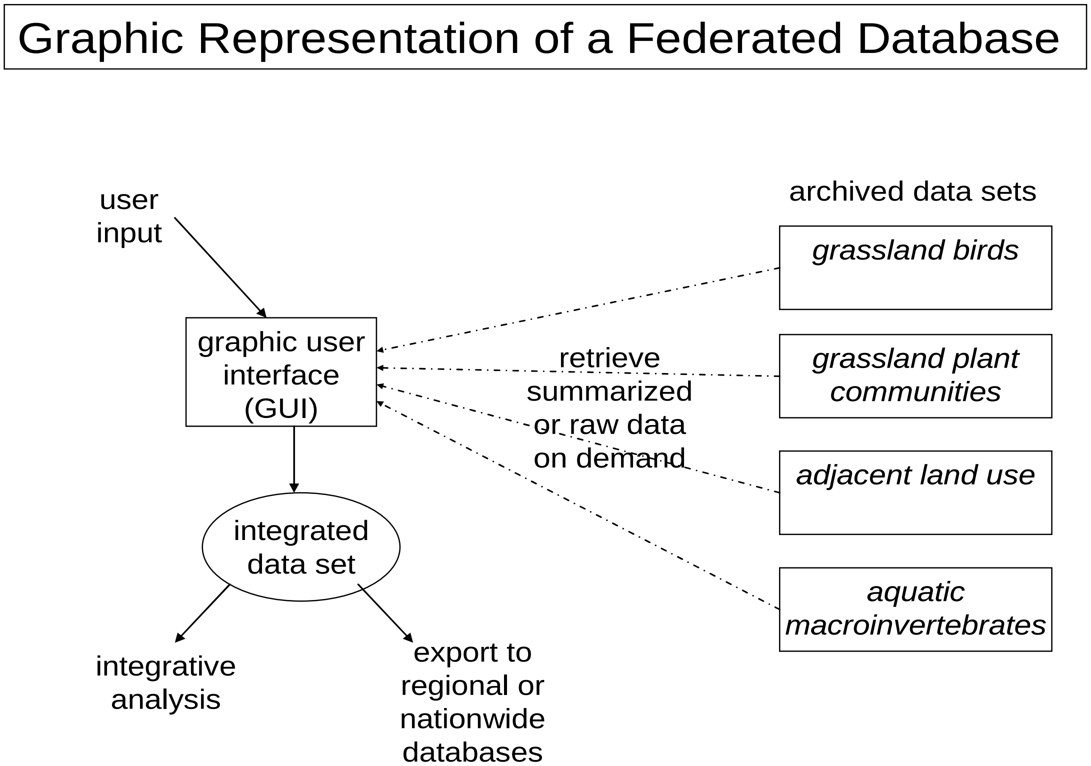


Figure 12. Federated model for database integration.

At the network level, the use of federated data model with shared core tables promotes tracking nationwide through the vital sign's framework (see Figure 12). Both federated model and vital signs framework provide the opportunity for qualitative assessment at the network or service-wide level. This is important as it provides glue to the network. Data management is the key to an integrated program, especially given the great diversity of resources in the Network.

7. Work Plan

In addition to documenting the data management accomplishments of HTLN, the following work plan directs future efforts by defining specific short-term, medium and long-term goals for each of the primary data management objectives. In order for this document to accurately reflect current goals and objectives, the Data Management Plan will be updated every 3-5 years.

Data infrastructure

Short term: **0-1 years**
- Complete database design and data entry forms for all monitoring projects (20% complete).
- Complete transition of legacy data into modern formats (70% complete).
- Finish description of project databases for inclusion in DMP appendices (20% complete).

Medium term: 1-3 years
- Standardize field trip reports.
- Convert all databases to the 2003 version of MS Access after 2003 is adopted as the service-wide standard.

Long term: 3-5 years
- Update databases for continued compatibility.

Data integration and exploration

Short term: 0-1 years
- Design prototype graphic user interface, queries and batch commands to integrate data (20% complete).
- Share data and reports through the I&M web page. Post entire data sets for downloading (with due consideration of security) and distribute reports as PDF files.
- Link spatial data and tabular data using the ArcGIS – Access geodatabase (30%).

Medium term: 1-3 years
- Automate routine report generation in MS Access (30% complete).
- Automate import and routine summary of ancillary data sets (e.g. precipitation and stream flow data for macroinvertebrate reporting).
- Integrate all data sets. Promote data integration among I&M Networks in the MWR through the use of standardized core data tables.
- Perform spatial analysis of plant community data.
- Perform spatial analysis of grassland bird data. Incorporate data with regional and national data sets.

Long term: 3-5 years
- Create a database to track resource management activities and integrate the database with other monitoring data.

Long-term data integrity and security

Short term: 0-1 years

- Update and automate tape backups of both the server and individual machines. Formalize a backup schedule (90% complete).
- Arrange for off-site data storage to guard against a catastrophic event. Schedule regular data shipments to the repository (90% complete).
- Implement data file organization scheme (90% complete).

Medium term: 1-3 years

- Create metadata for spatial data sets and archive data sets on the National Park Service GIS webpage.
- Log data sets into the Dataset Catalog and produce annual updates to NPSpecies and NatureBib.

Long term: 3-5 years

- Update Data Management Plan.

Literature Cited

Boetsch, J., M. DeBacker, P. Hughes, D. Peitz, L. Thomas, G. Wagner, and B. Witcher. 2000. A study plan to inventory vascular plants and vertebrates: Prairie Cluster and Heartland Network, National Park Service.

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., D. L. Borchers, L. Thomas. 2001. Introduction to Distance Sampling: Estimating abundance of biological populations. Oxford University Press. Oxford. 448 pp.

Carlin, B.P. and T.A. Louis. 2000. Bayes and Empirical Bayes Methods for Data Analysis, 2nd edition. 2000. Chapman and Hall/CRC. Boca Raton, FL. 419 pp.

Chatfield, C. 2004. The analysis of time series, an introduction, 6th edition. CRC Press. Boca Raton. FL. 333 pp.

Central Alaska Network Data Management Plan. 2004. Draft Chapter Chapter 2.

Cohen, J. 1988. Statistical power analysis for the behavioral sciences. Lawrence Erlbaum Associates Publishing. Hillsdale, NJ. 567 pp.

DeBacker, M.D., B. Witcher, and J.R. Boetsch. 2002. Data Management Plan: Prairie Cluster Prototype Long-Term Ecological Monitoring Program.

Fancy, S. 2004. NPS Inventory & Monitoring – Monitoring Natural Resources in Our National Parks. http://science.nature.nps.gov/im/monitor/index.htm

Hansen A. and D. Gryskiewicz. 2003. Interactions between Heartland National Parks and Surrounding Land Use Change: Development of Conceptual Models and Indicators for Monitoring. Final Report to the National Park Service Heartland Network.

Magurran, A.E. Ecological Diversity and Its Measurement. 1988. Princeton University Press. Princeton, NJ. 179 pp.

National Park Service. 1998. Channel Islands National Park, Data Management Protocol.

NCSS Statistical Software. 2004. Statistical Analysis and Graphics Software. Copyright 2003, Jerry Hintze. WWW.NCSS.COM

Northeast Coastal and Barrier Network Data Management Plan, Draft Chapter 9. 2004.

Northern Colorado Plateau Inventory and Monitoring Network Data Management Plan. Draft Plan. 2003.

Oakley, K. L., L. P. Thomas, and S. G. Fancy. 2003. Guidelines for long-term monitoring protocols. Wildl. Soc. Bull. 31(4):1000-1003.

Peitz, D.G., S.G. Fancy, L.P. Thomas, G.A. Rowell, and M.D. DeBacker. 2003. Bird Monitoring Protocol for Agate Fossil Beds National Monument, Nebraska and Tallgrass Prairie National Preserve, Kansas. Prairie Cluster Prototype Monitoring Program, National Park Service, U.S. Department of the Interior.

Peitz, D.G. and G.A. Rowell. 2004. Fish Community Monitoring in Prairie Park Streams with Emphasis on Topeka Shiner (*Notropis topeka*). Prairie Cluster Prototype Monitoring Program, National Park Service, U.S. Department of the Interior.

Pielou, E.C. 1969. An introduction to mathematical ecology. John Wiley and Sons, New York, New York. 286pp.

Shannon, C.E. 1949. The mathematical theory of communication. University of Illinois Press, Urbana, Illinois. 177 pp.

Snedecor, G.W. and W.G. Cochran. 1980. Statistical Methods, 7th edition. Iowa State University Press. Ames, Iowa. 507 pp.

Sokal, R.R. and F.J. Rohlf. 1995. Biometry, 3rd edition. W.H. Freeman and Co. New York. 887 pp.

Tessler, Steven & Joe Gregson. 1997. Draft Data Management Protocol. National Park Service.

Thomas, L. and C. Krebs. 1997. A review of statistical power analysis software. ESA Bulletin.

USAEC. 1999. US Army Environmental Center ITAM Technical Reference Manual: Ecological Monitoring on Army Lands. US Army Environmental Center, Aberdeen Proving Ground.

Appendix 1.

Standard Operating Procedure (SOP)
for Prelimary Data Analysis of Vital Signs Monitoring Data

Statistical Overview

Two factors permeate all data analysis and place constraints on subsequent biological interpretation of field data and management recommendations derived from such interpretation. Firstly, field data are samples and statistics derived from them are estimates of ‘real’ but unknown population dimensions or relationships. Secondly, in evaluating comparable statistics from different populations using a statistical test, not only must the significance level of the test be chosen with care, but the power of the test must be understood. In essence, it is necessary to test not only whether the possibility that the observed difference between samples could be due to chance, but also to understand the ability of the test to detect a difference between the samples given that a difference between populations actually exists. Thus, the guiding approach towards data analysis and interpretation of results must utilize appropriate methods and these will include tests of statistical significance together with associated power analyses. This approach is discussed in more detail below.

Data analysis starts by simply asking “What are the questions of interest?” Do they involve differences or correlations? If differences, are the tests between distributions, means, or variances? If correlations, do the data vary (correlate) with another variable or is one a function of another (regression). After the question of interest has been determined and the scale of measurement identified as either discrete (nominal or ordinal) or continuous (interval or ratio), exploratory data analysis can begin.

Because many statistical tests rely on certain assumptions about the data (i.e. type, distribution, balanced design, homogenous errors) all analyses must follow accepted procedures readily available in well-known statistical textbooks (e.g., Sokal & Rohlf, 1995). As most of the statistical tests our network employs are models of similar designs (ANOVA, RMANOVA, Regression), preliminary exploratory analyses consists of testing data for normality and homoscedasticity (homogeneity of variances).

Number Cruncher Statistical Systems (NCSS 2004) is used by our network to test for data normality using simple descriptive statistics. If raw data are not normally distributed, they commonly can be transformed into a useable normal format. The existence of homoscedasticity can be tested using Bartlett’s test (Sokal & Rohlf, 1995). Once model assumptions are met, analysis via SAS, NCSS, or other software proceeds.

Because statistical tests between say two populations or for the existence of a ‘treatment’ effect rely on samples of the populations under study, the results of the tests are based on probabilities that the data represent the ‘real’ condition of the population parameter-or statistical significance of the treatment effect (Snedecor and Cochran 1980). Two types of error are possible with such tests. If the statistical results are interpreted as indicating that a difference exists or a change has occurred with a probability of α and a 'real' difference or change has indeed occurred, the correct decision was made (no error). Alternatively, if there were no 'real' difference and yet a false difference was detected, this outcome is called a Type I error (also termed a 'false-change' error by some [USAEC, 1999])and will occur with a probability of α. The converse of a Type I error is a Type II error (a 'missed change' error in USAEC alternative terminology), when a test fails to determine that a 'real' difference exists. It occurs with a probability β. The relationships between the 'real' population condition, change detection, and statistical error are shown in Table A.1.1.

Table A.1.1. Types of error associated with using statistical tests.

	The 'Real' Population Condition	
Statistical Test Result	Difference exists	No difference exists
Difference detected	No Error **(**1-β)	Type I Error (α)
No difference detected	Type II Error (β)	No Error **(**1-α)

The significance to management of these two types of error is emphasized by their alternative names (USAEC, 1999). If a type I (false change error) is made, resources may be devoted to solve a problem that is a statistical artifact. If a type II (missed change error) is committed, a significant problem, which potentially merits action, will be overlooked. Tests for significant differences using α have been employed for decades. A practical understanding of the importance β and the topic of type II errors (Cohen, 1988) has been explored only more recently and their use in ecological work is a feature of the past decade (Thomas and Krebs 1997, USAEC, 1999).

Power analysis, where power is (1-β), examines the factors involved in type II (missed change) errors. The topic is well reviewed by Thomas and Krebs, 1997and USAEC, 1999. The power of a test is related to five factors: the choice of β, the choice of α, variation in the data set, sample size, and the minimum detectable change or 'effect size.' Knowledge of any four of these factors enables the fifth to be calculated.

Power analyses can be used *a priori* (i.e. in sample design) provided pilot studies have yielded guidance on data variability. In this setting, after acceptable levels of α and β are chosen, a decision is made on the minimum difference or change that has biological significance. This latter value is used for the minimum detectable change and allows calculation of the needed sample size.

Power analysis may also be employed *a posteriori*, to determine if protocols are adequate. The consequences of various levels of power can be explored. How do such changes in level of power affect the minimum detectable change or 'effect size' and thus the ability to determine differences between populations (Cohen, 1988). Clearly, the minimum detectable change should closely correspond to a difference of biological significance (USAEC, 1999). If, at an acceptable level of power, the minimum detectable change is smaller than the difference of biological significance, the test is strong. Conversely, if the minimum detectable difference at a given power is bigger than the difference of biological interest, the study design is weak.

In summary, whether or not the results of the statistical test can be considered 'true' or reliable depends on the power of the statistical test. If a test is significant and has high power then the results can be deemed valid and confidence attached to the conclusions. Conversely, if a test is significant but has low power no reliable conclusions can be drawn from the data. Failure to use power analysis in the interpretation of statistical tests for change detection may easily lead to faulty management recommendations as noted above, or lead to squandering of field resources when the minimum detectable change is well below the threshold of anything of biological significance.

Example of a descriptive statistics report produced from PASS.

Descriptive Statistics Report
Page/Date/Time 1 5/20/2004 2:18:26 PM
Database

Summary Section of Y when Treatment=A

Count	Mean	Standard Deviation	Standard Error	Minimum	Maximum	Range
60	90.18035	17.0183	2.197052	52	124.7117	72.71175

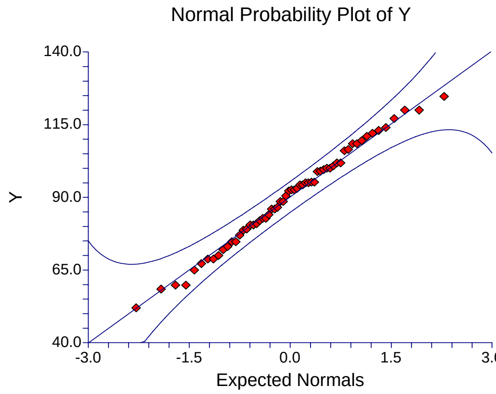
Mean-Deviation Section of Y when Treatment=A

Parameter	|X-Mean|	|X-Median|	(X-Mean)^2	(X-Mean)^3	(X-Mean)^4
Average	13.89141	13.81442	284.7953	-536.5244	193980.4
Std Error	1.323036		43.37281	940.8614	49489.89

Normality Test Section of Y when Treatment=A

Test Name	Test Value	Prob Level	10% Critical Value	5% Critical Value	Decision (5%)
Shapiro-Wilk W	0.9885094	0.844818			Accept
Anderson-Darling	0.1626961	0.945145			Accept
Martinez-Iglewicz	0.9549347		1.080159	1.123554	Accept
Kolmogorov-Smirnov	4.672739E-02		0.104	0.114	Accept
D'Agostino Skewness	-0.3872	0.698628	1.645	1.960	Accept
D'Agostino Kurtosis	-1.0292	0.303400	1.645	1.960	Accept
D'Agostino Omnibus	1.2091	0.546322	4.605	5.991	Accept

Plots Section of Y when Treatment=A



Example of Analysis of Variance Report (NCSS).

Expected Mean Squares Section

Source Term	DF	Term Fixed?	Denominator Term	Expected Mean Square
A: Treatment	2	Yes	S	S+sA
S	177	No		S

Note: Expected Mean Squares are for the balanced cell-frequency case.

Analysis of Variance Table

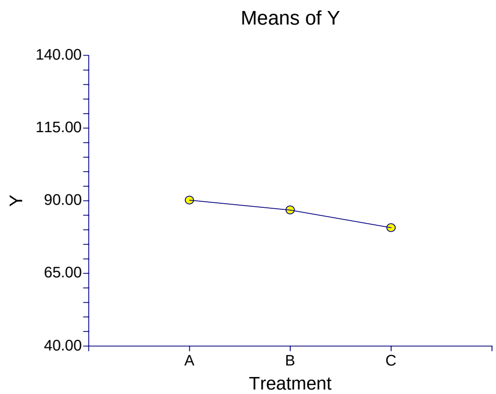
Source Term	DF	Sum of Squares	Mean Square	F-Ratio	Prob Level	Power (Alpha=0.05)
A: Treatment	2	2763.917	1381.958	4.49	0.012585*	0.761737
S	177	54536.07	308.1134			
Total (Adjusted)	179	57299.99				
Total	180					

* Term significant at alpha = 0.05

Means and Effects Section

Term	Count	Mean	Standard Error	Effect
All	180	85.89377		85.89377
A: Treatment				
A	60	90.18035	2.266103	4.286583
B	60	86.79211	2.266103	0.8983431
C	60	80.70884	2.266103	-5.184926

Plots Section

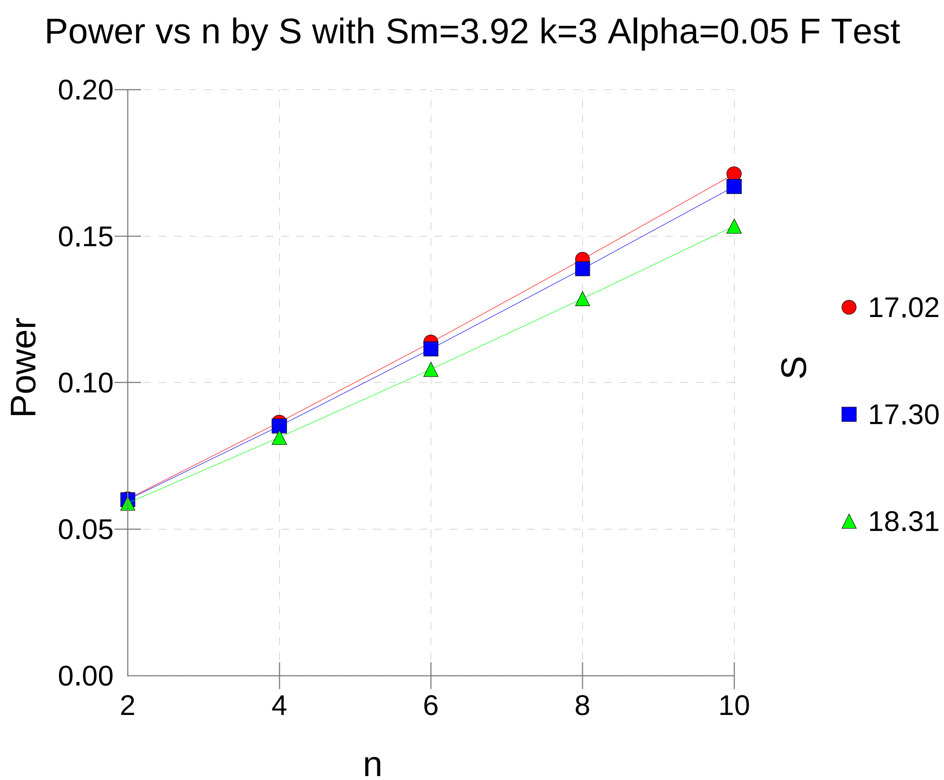


Abbreviated example of power analysis (PASS).

Power	Average n	k	Total N	Alpha	Beta	Std Dev of Means (Sm)	Standard Deviation (S)	Effect Size
0.06037	2.00	3	6	0.05000	0.93963	3.92	17.02	0.2302
0.08649	4.00	3	12	0.05000	0.91351	3.92	17.02	0.2302
0.11378	6.00	3	18	0.05000	0.88622	3.92	17.02	0.2302
0.14211	8.00	3	24	0.05000	0.85789	3.92	17.02	0.2302
0.17131	10.00	3	30	0.05000	0.82869	3.92	17.02	0.2302
0.06003	2.00	3	6	0.05000	0.93997	3.92	17.30	0.2264
0.08525	4.00	3	12	0.05000	0.91475	3.92	17.30	0.2264
0.11156	6.00	3	18	0.05000	0.88844	3.92	17.30	0.2264

Power is the probability of rejecting a false null hypothesis. It should be close to one. n is the average group sample size; k is the number of groups; Total N is the total sample size of all groups; Alpha is the probability of rejecting a true null hypothesis. It should be small; Beta is the probability of accepting a false null hypothesis. It should be small; Sm is the standard deviation of the group means under the alternative hypothesis; Standard deviation is the within group standard deviation; The Effect Size is the ratio of Sm to standard deviation.

Summary Statements
In a one-way ANOVA study, sample sizes of 2, 2, and 2 are obtained from the 3 groups whose means are to be compared. The total sample of 6 subjects achieves 6% power to detect differences among the means versus the alternative of equal means using an F test with a 0.05000 significance level. The size of the variation in the means is represented by their standard deviation which is 3.92. The common standard deviation within a group is assumed to be 17.02.



Literature Cited

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., D. L. Borchers, L. Thomas. 2001. Introduction to Distance Sampling: Estimating abundance of biological populations. Oxford University Press. Oxford. 448 pp.

Carlin, B.P. and T.A. Louis. 2000. Bayes and Empirical Bayes Methods for Data Analysis, 2nd edition. 2000. Chapman and Hall/CRC. Boca Raton, FL. 419 pp.

Chatfield, C. 2004. The analysis of time series, an introduction, 6th edition. CRC Press. Boca Raton. FL. 333 pp.

Cohen, J. 1988. Statistical power analysis for the behavioral sciences. Lawrence Erlbaum Associates Publishing. Hillsdale, NJ. 567 pp.

Magurran, A.E. Ecological Diversity and Its Measurement. 1988. Princeton University Press. Princeton, NJ. 179 pp.

Peitz, D.G., S.G. Fancy, L.P. Thomas, G.A. Rowell, and M.D. DeBacker. 2003. Bird Monitoring Protocol for Agate Fossil Beds National Monument, Nebraska and Tallgrass Prairie National Preserve, Kansas. Prairie Cluster Prototype Monitoring Program, National Park Service, U.S. Department of the Interior.

Peitz, D.G. and G.A. Rowell. 2004. Fish Community Monitoring in Prairie Park Streams with Emphasis on Topeka Shiner (*Notropis topeka*). Prairie Cluster Prototype Monitoring Program, National Park Service, U.S. Department of the Interior.

Pielou, E.C. 1969. An introduction to mathematical ecology. John Wiley and Sons, New York, New York. 286pp.

Shannon, C.E. 1949. The mathematical theory of communication. University of Illinois Press, Urbana, Illinois. 177 pp.

Snedecor, G.W. and W.G. Cochran. 1980. Statistical Methods, 7th edition. Iowa State University Press. Ames, Iowa. 507 pp.

Sokal, R.R. and F.J. Rohlf. 1995. Biometry, 3rd edition. W.H. Freeman and Co. New York. 887 pp.

Thomas, L. and C. Krebs. 1997. A review of statistical power analysis software. Ecological Society of America, Bulletin 78: 126-139.

USAEC. 1999. US Army Environmental Center ITAM Technical Reference Manual: Ecological Monitoring on Army Lands. US Army Environmental Center, Aberdeen Proving Ground.